University of Kentucky
UKnowledge

Theses and Dissertations--Education Science | College of Education

2017

TOWARD ASSESSMENT LEADERSHIP: STUDY OF ASSESSMENT PRACTICES AMONG SCHOOL AND CLASSROOM LEADERS

Carrie Elizabeth Eubank Morris
University of Kentucky, carrieemorris@gmail.com
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.462

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation

Eubank Morris, Carrie Elizabeth, "TOWARD ASSESSMENT LEADERSHIP: STUDY OF ASSESSMENT PRACTICES AMONG SCHOOL AND CLASSROOM LEADERS" (2017). *Theses and Dissertations--Education Science*. 30.
https://uknowledge.uky.edu/edsc_etds/30

This Doctoral Dissertation is brought to you for free and open access by the College of Education at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Education Science by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Carrie Elizabeth Eubank Morris, Student

Dr. Tricia Browne-Ferrigno, Major Professor

Dr. Justin Bathon, Director of Graduate Studies

TOWARD ASSESSMENT LEADERSHIP: STUDY OF ASSESSMENT PRACTICES AMONG SCHOOL AND CLASSROOM LEADERS

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Education

at the University of Kentucky

By
Carrie Elizabeth Eubank Morris

Lexington, Kentucky

Co-Directors: Dr. Tricia Browne-Ferrigno, Professor of Educational Leadership Studies
And Dr. Michael D. Toland, Associate Professor of Educational, School, and Counseling Psychology

Lexington, KY

2017

Copyright © Carrie Elizabeth Eubank Morris 2017

ABSTRACT OF DISSERTATION

TOWARD ASSESSMENT LEADERSHIP: STUDY OF ASSESSMENT PRACTICES AMONG SCHOOL AND CLASSROOM LEADERS

Traditionally, models of instructional leadership espouse data-informed decision making in response to student assessment outcomes as one of the core school leader behaviors. In recent years, rising expectations from accountability policies and related assessment practices have myriad implications for school districts, specifically in the areas of standards-driven reform, student assessment systems, and professional development models. As a result, demands on schools to collect and use student assessment data to inform curricular and instructional decisions has expanded. While principals are typically held responsible for school improvement efforts, more contemporary models of instructional leadership incorporate teachers as classroom-based leaders of assessment practices in forums such as professional learning communities.

School and classroom assessment leaders engage in behaviors such as (a) identifying an assessment vision, (b) fostering group goals, (c) providing a model of data-informed decision making, (d) promoting teacher job-embedded professional learning experiences, (e) evaluating instructional practices with specific feedback, and (f) strategically aligning resources to school improvement goals. Unfortunately, school districts face many challenges with assessment leadership due to barriers in beliefs about assessments, time with and access to tools and training, and knowledge and skills about how to operationalize effective assessment practices that yield positive student outcomes.

The purpose of this study was to explore assessment leadership as a construct among P-12 school and classroom leaders in one large district in Florida. Data were collected using an Internet-based survey constructed from existing qualitative and quantitative measures of key components of assessment leadership established in the literature. A series of descriptive and inferential analyses were conducted to (a) explore the factor structure of the instrument and (b) evaluate the influence of assessment learning experiences, beliefs, and knowledge on assessment practices. Relationships among variables were examined when considering moderating variables for school role (i.e., school-level administrator or classroom teacher as professional learning communities facilitator) and school type (elementary or secondary). Limitations were discussed to inform future research in this critical area of school improvement.

KEYWORDS: School Accountability, Assessment Leadership, Assessment Literacy, Classroom Assessment Practices, Professional Learning Communities

Carrie Elizabeth Eubank Morris

December 8, 2017
Date

TOWARD ASSESSMENT LEADERSHIP: STUDY OF ASSESSMENT PRACTICES AMONG SCHOOL AND CLASSROOM LEADERS

By

Carrie Elizabeth Eubank Morris

Dr. Tricia Browne-Ferrigno
Co-Director of Dissertation

Dr. Michael D. Toland
Co-Director of Dissertation

Dr. Justin Bathon
Director of Graduate Studies

December 8. 2017
Date

This paper is dedicated to four greats who came before me: Earl Glover, Gladys Glover, Dr. William Procter Eubank, Sr., and Elizabeth Eubank. These individuals laid the foundations for our family, for what my parents would eventually instill in me: Be Kind, Be Thankful, and Never Give Up.

ACKNOWLEDGEMENTS

The following dissertation, while an individual accomplishment, benefited from the insights and direction of several people. First, I am grateful to Professor Tricia Browne-Ferrigno, my advisor and chair of my dissertation committee. From the start of my studies at the University of Kentucky, she motivated me to pursue this goal with relentless fervor and unyielding spirit. I recognize I have been a non-traditional type student: I chose to embark on this journey in the middle of my career, which meant balancing with each breath the role of student with the roles of mom, daughter, partner, friend, and professional. She understood. And when my life took an unexpected turn, she extended the just-right dose of compassion and tenacity that caused me to persist. Most importantly, she refused to let me decide that the journey had ended without it being finished. I submit this manuscript knowing that, even when there was silence on my part, Tricia upheld her commitment to me, to meet me at the end of this path.

Next, I want to acknowledge Dr. Michael Toland, my dissertation co-chair. I feel I may have landed the best pair of co-chairs imaginable. He offered me the expertise I needed to handle every statistical procedure, known and unknown, but necessary to learn to accomplish this dissertation with integrity. He was willing to dedicate time to me while also accommodating my demanding work schedule and distance situation. He shared with ease his skills and resources, guiding me to establish understanding and make statistically sound and informed decisions. Most importantly, he challenged me. I am a proponent of the sentiment that nothing worth having comes easily. This experience has been no different. Michael helped me meet the challenge at its most difficult moments.

Also, to all the members of my committee, Professor Lars Bjork, Dr. Wayne Lewis, and Professor Tom Guskey: each of them in differing ways inspired me to approach education with steadfast curiosity and unwavering commitment to continuous improvement. They, along with Tricia and Michael, are teachers. They are leaders. They are innovators. I aspire to leave a mark that approximates their collective influence.

Finally, I dedicate this accomplishment to those individuals that move me where the mind meets the heart. I am fortunate in this life to call my family people that stand with each other, grow with each other, stay connected with each other, and still like to vacation with each other. Each one is varied in their role and respective influence on me. But, in their own ways, direct and indirect, they have encouraged and sustained me throughout this journey. It would take me another dissertation to articulate the many ways, so I state their names as representative of the profound time and space they know we share in this life: Dad, Mom, Rebecca, Nick, Jillian, Everly, LeeAnn, Bob, Chris, Alex, Katie, Moey, Joey, Jenny, Mark, Cortney, Matt, Diane, Ralph, Jim, Karen, Christopher, and Cam. I am equally as fortunate to have gained friends and colleagues I consider like family. In moments they may or may not have realized, they gifted me with affection, challenge, reassurance, amusement, and so much more: Ellen, Erica, Jenine, Michelle, Heidi, Steve, George, Vicki, Alyssa, Jenna, Amber, Daniela, Blaire, Melissa, Holly, Celisa, Amy, and Judy. To those who have served as my supervisors, I own my passions for learning, teaching, and innovating: Amelia, Mike, and Julene. My final acknowledgements are reserved for my partner Chris and my daughter Elizabeth. Chris meets me exactly where I need with laughter and love. For Elizabeth, this work is my best illustration to date of my hope for her life: be kind, be thankful, and never give up.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Educational movements in school accountability, accompanied by standards-based and assessment reforms, have led to heightened expectations for states and school districts to demonstrate improved academic outcomes and graduation rates for all students (Council of Chief State School Officers [CCSSO], 2010; National Governors Association [NGA], 2010; United States Department of Education [USDE], 2002, 2004, 2009, 2015). In response, schools have evolved instructional models to systematically incorporate standards-based curricula, student-assessment systems, and research-based interventions (Clark, 2011; Means, Padilla, DeBarger, & Bakia, 2009). These movements represent a departure from traditional approaches to learning and teaching, including how schools measure and respond to student-learning outcomes. At the forefront of reform are principals and teachers collecting and using standards-based assessment data to inform and monitor instructional decisions in the classroom as well as for school improvement plans (Coburn & Talbert, 2006; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Shen, Cooley, Reeves, Burt, Ryan, Rainey, & Yuan, 2010).

Evolution of School Accountability

Accountability policies linked to large-scale assessments have existed for nearly 50 years. The initial legislation, *Elementary and Secondary Education Act* (ESEA) (USDE, 1965), called for schools to evaluate the effectiveness of educational programs with a primary focus on students from low socioeconomic areas. President Johnson was intent on tackling poverty through education by providing underprivileged students with improved and equitable educational opportunities. The hope was to close

achievement gaps for all students, regardless of their socioeconomic status. State and district responses to the new legislation included the creation of competency assessments (i.e., measures of minimum expectations of academic skills). Florida was one of the first states to implement such assessments and to use the outcomes as readiness measures for graduation. Teacher names were attached to their students' response sheets for English and mathematics examinations, which were provided to the state department of education (Linn, 2000). Even though the assessments were designed to gauge student performance, they were not considered robust enough to capture the comprehensive knowledge and skills expected of all students, including those from disaggregated groups. Even so, this seminal work set the foundation for future developments in standards-based assessments.

Introduction of Assessment-Driven Policies

In the 1980s, the National Commission on Excellence published the report, *A Nation at Risk* (USDE, 1983), demarking the first movement toward heightened expectations for states and districts to demonstrate improved student outcomes (Hanushek & Raymond, 2005). The report summarized results of available high-stakes assessments from prior years, including the state competency assessments and the Scholastic Aptitude Test (SAT). Evidence of the decline on these assessments nationwide prompted numerous recommendations from the federal government to state and district leaders and ultimately underscored the systematic inclusion of comprehensive, high-stakes assessments as critical elements of educational reform.

Specifically, the report challenged legislators and educators to (a) expand required content for graduation, (b) advance expectations for student achievement based on college admission standards, (c) increase school-day time in instruction, (d) certify

expectations for competencies in teaching practices, and (e) embrace fiscal and civil leadership responsibilities at the federal government level (USDE, 1983). In response, all but a few states established accountability policies requiring districts to adopt curriculum standards and develop academic assessments to measure standards mastery (Hanushek & Raymond, 2005). Even after expanded implementation of student assessments linked to specific academic standards, it was not until the 1990s, when standard-based reform truly mobilized, that accountability policies evolved to a higher level.

In 1994, passage of the *Goals 2000: Educate American Act* increased the federal government's role in public education with heightened expectations for states and districts, prompting shifts in standards-based reform. *Goals 2000* required educators to concentrate accountability policies in new areas, specifically Kindergarten readiness, increased graduation rates, improved literacy, violence prevention, greater parental involvement, and enhanced teacher professionalism (USDE, 1994). With this legislation came additional specifications for high-stakes tests, including expectations for student achievement across all content areas at designated grade levels. Federal funds were allocated in two key areas: (1) provide technological resources to schools and (2) expand the professional development activities of teachers and leaders. Although none of the goals were achieved by year 2000, implications of the legislation were significant as states and districts dedicated resources to improve student outcomes with formalized curriculum standards aligned to standardized assessment measures.

Expansion of Standards-Based Movements

During the first decade of the 21st century, educators experienced considerable acceleration in educational policies and practices. Movement toward accountability for

all increased pressures for schools to adopt common academic standards and demonstrate student-learning outcomes linked to district- and school-improvement plans. The 2002 reauthorization of the ESEA, more commonly known as *No Child Left Behind* (NCLB), required all states and districts to establish academic standards and administer assessments as annual measures of progress toward the standards. The enactment of NCLB, coupled with the reauthorization of the *Individuals with Disabilities Education Act* (IDEA) (USDE, 2004, 2007), marked a significant shift in standards-based assessment reform—from schools merely providing students *access to academic content* to schools also *demonstrating effectiveness* through student learning-outcomes.

No Child Left Behind. Prior to the enactment of NCLB, states had been responsible for establishing parameters for student achievement. NCLB mandated that states and districts expand previous accountability policies and high-stakes assessment systems to incorporate school-grading formulas as indicators of school effectiveness (Hursh, 2007). The formulas included specific expectations for disaggregated student groups, such as students with disabilities, students from diverse ethnic groups, English language learners, and students from economically disadvantaged backgrounds. In addition, for the first time in the evolution of accountability policies, teacher quality was introduced as a component of school effectiveness. States and districts expanded teacher certification requirements to include definitions of highly effective teachers and teaching practices (Youngs, 2013). Furthermore, these new federal policies mandated states exercise consequences such as offering families alternative school placements and vouchers for districts and schools who did not meet expectations. States also were

subject to sanctions including loss of funding and removing personnel in cases of repeated lack of progress toward annual goals (Darling-Hammond, 2004; Hursh, 2007).

Race to the Top. Although the ESEA was scheduled for reauthorization in 2012, it was not revised by its renewal date. In the absence of new legislation, President Obama developed a blueprint for educational reform that outlined specifications for extended emphases on standards-based instruction, assessment systems, and teacher quality (USDE, 2010). Federal funds, in the form of Race to the Top grants (USDE, 2009), were awarded to states that produced updated and viable plans to support continued improvements in effective instructional programs and assessment systems. The minimum requirements for funding involved (a) adoption of comprehensive academic standards, (b) enhancements to statewide assessments with an emphasis on measuring student growth, and (c) implementation of teacher evaluation systems that incorporated student outcomes as a core component of evidencing teacher effectiveness (Clark, 2011; Youngs, 2013). As incentives, states were not only provided with federal funding but also given waivers from the NCLB-mandated sanctions if they failed to meet the 2014 deadlines for adequate yearly student progress (USDE, 2012). Forty-two (42) of the 50 states took advantage of this option to evade sanctions extending from NCLB.

Every Student Succeeds Act. In January 2015, Secretary of Education Duncan called for a renewed commitment to the main tenets of the original ESEA, enacted under President Johnson's administration in 1965 (USDE, 2015). Shortly afterward, President Obama signed into legislation the reauthorization of ESEA, entitled *Every Student Succeeds Act* (ESSA) (USDE, 2015). This federal legislation built upon prior years of accountability policies with continued focus on (a) providing equitable educational

opportunities for all students, (b) assessing regularly student-learning outcomes toward achieving common academic standards, and (c) implementing high-quality educator effectiveness programs with a shift toward more state control in these areas. New language included expectations for college and career readiness, innovation in instructional and technological practices, early intervention including pre-school experiences, and wraparound support systems for schools located in at-risk communities.

In 2016, ESSA was put into motion by the federal government with requests for school districts to submit accountability plans for the 2017-18 school year. With the new administration, states were provided with more flexibility than in recent year accountability policy implementations. For example, states could set their own goals with the continued expectation that student groups below learning targets were identified for closing the achievement gap and increasing graduation rates (Editorial Projects in Education Research Center, 2016). In the lowest performing schools, school districts were required to implement interventions and measure student-learning outcomes. However, many of the mandates present in previous policies such as teacher evaluation and adoption of common core curriculum standards, were relaxed. Even though the recent enactment of ESSA represents a new era of accountability, the focus on student achievement linked to standards-based assessment remains central to educational reform.

Implications of Educational Policies on Practices

Since 1965, accountability policies have underscored educational reform in the areas of curriculum, instruction, and assessment as well as for leadership behaviors. At the district level, leaders responded by directing school-improvement efforts toward (a) defining academic standards, (b) innovating curricular and instructional practices, (c)

developing assessment monitoring systems, and (d) designing professional growth systems that promote instructional leadership among principals (Clifford & Mason, 2013; Youngs, 2013). At the school level, principals' roles and responsibilities expanded beyond the traditional management of building facilities and human resources to implementing school district initiatives in these areas to improve learning and teaching.

Specifically, principal activities evolved to include directly coaching and modeling teaching practices in the classroom, arranging instructional planning and professional learning opportunities for teachers, and collecting and reviewing data to guide school-level instructional decisions (Hattie, 2009; Sebastian & Allensworth, 2012; Stiggins & Duke, 2008). Although these tasks were considered critical instructional leadership behaviors, oftentimes they far exceeded what was possible for one person to complete alone (Sanders & Kearney, 2008). As a result, many principals embraced leadership models where roles and responsibilities were shared among teacher leaders (Gedik & Belibas, 2015; Renihan & Noonan, 2012). The distribution of roles and responsibilities promotes a culture of inquiry and collaboration among teaching professionals to meet the mounting expectations extending from accountability policies.

Curriculum and instruction updates. Historically, districts have provided schools with guidance on curricular programs and instructional strategies to support teachers in effectively planning and delivering instruction as well as in assessing student-learning outcomes. While traditional approaches primarily have been rooted in content knowledge and pedagogy, contemporary approaches to curriculum and instruction have been fueled by standards-based and assessment reforms (Sanders & Kearney, 2008). As a result, school districts have been faced with re-conceptualizing curriculum programs to

incorporate academic content and innovative instructional practices linked to grade-level standards and measured by statewide and local assessments (Stiggins, 2005).

In the wake of curricular and instructional reform, common academic standards (CCSSO, 2010), coupled with emphases on 21st learning competencies, social-emotional learning skills, and instructional technology systems (NGA, 2010), prompted many districts to develop innovative and differentiated approaches that addressed the complexities of the enhanced academic standards (Meyers & Murphy, 2007). These curricula included new content materials, technological resources, and standards-based instructional tools and practices (Elfers, 2008; Fisher, 2005; Hattie, 2012; Loeb, Knapp, & Elfers, 2008). In recent years, enactment of ESSA with language pertaining to college and career readiness and innovative practices led to continued exploration of advanced instructional and technical practices in districts and schools.

While new standards-based curricular materials and tools are intended to operationalize educational reform at the classroom level, many teachers are not prepared for implementing them (Clifford & Mason, 2013; Ulmer, 2002). As a result, principals are tasked with not only communicating district initiatives but also leading, supporting, and monitoring teachers' curriculum and instruction efforts—as well as allocating time and resources and arranging for targeted professional development and feedback in these areas (Deneen & Brown, 2016; Renihan & Noonan, 2012; Supovitz & Klein, 2003). To lead the efforts, principals must develop their own knowledge of and skills in standards-based approaches so that teachers have effective guidance and fidelity of implementation (Sanders & Kearney, 2008; Noonan & Renihan, 2006, 2010; Stiggins & Duke, 2008).

Student and teacher monitoring systems. In conjunction with new curricula and instructional efforts, districts and schools also have responded to accountability policies by developing comprehensive student assessment systems that integrate various formative and summative assessment tools beyond traditional grading practices (Gallagher, Means, & Padilla, 2008; Kingston & Nash, 2011; Stiggins & Duke, 2008). Classroom assessments are intended to measure progress toward academic standards and require teachers to adopt or create tools that are reflective of student-learning outcomes during instruction as well as after instruction (Means, Padilla, DeBarger, & Bakia, 2009; Young & Kim, 2010). These assessments are intended to be used formatively to inform teachers about student progress and leaders about learning trends (Stiggins, 2001).

Specifically, student assessment systems have expanded from grades alone to incorporate other methods such as classroom formative assessments, grade-level common assessments, and interim assessments that occur at quarter or semester intervals (Goertz, Olah, & Riggan, 2009). Assessments are designed to measure student progress toward academic standards throughout the school year and provide teachers with data to evaluate, plan, and differentiate instruction based on student needs in the moment (Black & Wiliam, 1998; Guskey, 2003; Stiggins, 2005). Such shifts in assessment practices influenced principal and teacher practices to incorporate collecting, analyzing, and using data to make school and classroom decisions (Coburn & Talbert, 2006; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Means, Padilla, & Gallagher, 2010; Noonan & Renihan, 2006; Shen et al., 2010). To accomplish these tasks, principals have arranged time and resources in collaborative sessions during the school weeks (Halverson, Grigg, Prichett, & Thomas, 2007; Loeb et al., 2008; Militello, Schweid, & Sireci, 2010).

In addition to creating and implementing student assessment systems that measure student-learning outcomes, accountability policies have prompted districts to establish systems for principals to evaluate teacher quality and effectiveness (Jenkins, 2009; Leithwood, Harris, & Hopkins, 2008; Youngs, 2013). In response, districts have adopted instructional monitoring approaches such as learning walkthroughs, instructional rounds, and data reviews to measure implementation toward school-improvement efforts (Clifford & Mason, 2013). At the school level, principal and teacher leaders increasingly are involved in activities such as conducting classroom observations, providing feedback to teachers, and monitoring progress toward professional development plans with the intention of improving teachers' instructional and assessment practices (Youngs, 2013). Extensive time, resources, and expertise are required by schools to develop and implement monitoring systems for students and teachers—with a level of integrity that yields positive student learning-outcomes and school-improvement goals.

Professional development opportunities. Shifts in accountability policies and practices require enhanced knowledge and skills for teachers and leaders to meet effectively the heightened demands (Noonan & Renihan, 2006, 2010; Stiggins & Duke, 2008). Unfortunately, most teacher and principal preparation programs target traditional aspects of learning, teaching, and leading that fail to incorporate extensive training in assessment-related skills development (Coggshall, Bivona, & Rechsley, 2012; Greenberg, McKee, &Walsh, 2013). Only in recent years have undergraduate and graduate programs incorporated coursework on standards-based instruction and formative assessment practices for teachers and created learning-centered professional environments for principals (Bryan & Simone, 2012; Greenberg et al., 2013; Sanders & Kearney, 2008).

To support principals and teachers in assessment efforts, many districts have refined professional development to incorporate job-embedded learning experiences such as professional learning communities (PLCs). PLCs systematically integrate professional development centered on problems of practice and reform initiatives into the day-to-day teacher experience (Danielson, 2009; Dufour & Eaker, 1998; Jenkins, 2009; Leithwood, Harris, & Hopkins, 2008). They promote professional conversations to plan instruction, create assessments, evaluate student outcomes, and share strategies for cultivating effective teaching practices (Danielson, 2009; Dufour & Eaker, 1998). PLCs have been shown to enhance learning for educational professionals, resulting in improved classroom practices and positive effects on student-learning outcomes (Coggshall et al., 2012; Hattie 2009; Robinson, Lloyd, & Rowe, 2008; Valentine & Prater, 2011), particularly when principals ensure teachers are provided with adequate and dedicated time and resources to these activities (Witziers, Bosker, & Kruger, 2003; Renihan & Noonan, 2012). As a result, PLCs have become a viable forum for learning and teaching work.

Statement of the Problem

Traditionally, data-informed decision making has been considered a core component of effective school leadership practices (Boudett, City, & Murnane, 2010; Fullan, 2001; Green, 2010; Hallinger, 2001, 2011; Horng & Loeb, 2000; Leithwood & Louis, 2012; Waters, Marzano, & McNulty, 2003). As discussed, recent shifts in accountability policies have intensified pressures for schools to innovate student assessment and data-use practices to ensure all students achieve the desired learning targets (Black, Harrison, Lee, Marshall, & Wiliam, 2004). Unfortunately, many educators report lacking the requisite skills to design assessments, analyze student data,

and implement effective protocols at a level that successfully impacts student-learning outcomes (Buck & Trauth-Nare, 2009; Ingram, Louis & Schroeder, 2004; Volante, 2010; Young & Kim, 2010). They also report gaps in time and resources for such activities.

Principals as instructional leaders serve a critical role in increasing teacher effectiveness (Diamond, Burch, Hallett, Jita, & Zoltners, 2002; Leithwood, Steinbach, & Jantzi, 2002; Marks & Nance, 2007) and improving student achievement (O'Day, 2002; Witziers, Bosker, & Kruger, 2003). Moreover, they are essential to ensuring schools embrace and implement reform. Central to their role is leading the efforts on student assessment systems using student data to inform school decisions. Principals as assessment leaders employ strategies such as creating a vision, fostering group goals, modeling effective practices, promoting teacher learning and development, planning curriculum, evaluating teaching practices with specific feedback, and strategically aligning resources to instructional goals (Green, 2010; Hattie 2009; Noonan & Renihan, 2006; Robinson et al., 2008; Valentine & Prater, 2011). However, for principals to embody the assessment-leader role successfully, they must first develop key competencies in assessment knowledge, appreciations (i.e., beliefs), and skills (Earl & Fullan, 2003; Noonan & Renihan, 2006; Popham, 2010; Stiggins & Duke, 2008).

Unfortunately, research in assessment leadership is limited. To date, studies have focused more narrowly on assessment literacy levels of principals and teachers than on assessment leadership as a construct (Alkharusi et al., 2011; Beziat & Coleman, 2015 Brookhart, 2001; Campbell, Murphy, & Holt, 2002; Deluca, LaPoint-McEwan, & Luhanga, 2016; Plake, Impara, & Fager, 1993). Even so, the importance of assessment leadership practices in assessment reform efforts is evident (Noonan & Renihan, 2006,

2010; Young & Kim, 2010). While there is agreement that assessment leadership practices can contribute to school success, there are gaps. First, there is a paucity of research aimed at defining and measuring key components of assessment leadership and its relative influence on classroom practices (Brookhart, 2001). Second, principals face many barriers to engaging in assessment leadership practices, such as time, access to data systems, beliefs about assessments, knowledge and skills in using data appropriately, and targeted professional development in effective assessment practices (Clifford & Mason, 2013; Volante & Cherubini, 2011). Thus, it is important to deepen the investigation of the underlying factors of assessment leadership to evolve research and inform practice.

Purpose and Significance of this Study

The purpose of this study was to investigate assessment leadership as a construct to gain more information about the critical components that contribute to successful implementation of assessment leadership practices in schools. The conceptual framework for assessment leadership incorporates specific leadership practices that have been identified in research and practice, including (a) setting a vision, (b) designing and using data systems, (c) leading data discussions, (d) developing teachers in effective assessment practices, and (e) self-reflecting on assessment practices (Noonan & Renihan, 2006). The key variables in this study were derived from factors hypothesized to contribute to assessment leadership practices: beliefs about assessments, assessment knowledge (defined as assessment literacy), and access to professional development in effective assessment practices and data use (Burke & Wang, 2010; Coburn & Talbert, 2006; Deenen & Brown, 2016; Gallagher et al, 2008; Kerr et al., 2006).

Even though instructional leadership activities typically have been associated with principals, this study sought to measure the assessment leadership practices performed by both school and classroom leaders. Within newer models of instructional leadership, it is argued that both principals and teachers are critical to school improvement efforts aimed at advancing student-learning outcomes (Hallinger & Heck, 2010; Lewis, Leithwood, Wahlstrom, & Anderson, 2010). As school leaders, principals are expected not only to coach and model effective practices in the classroom but also to distribute leader roles and responsibilities to teachers (Collins, 2001; Green, 2010). For this reason, principals, assistant principals, and classroom teachers who serve as teacher leaders in the role of PLC facilitator were included as part of the target population in this study.

This study is significant because it extends previous research in instructional leadership to include assessment leadership practices. Moreover, it seeks to investigate variables that influence the development of school and classroom leaders as assessment leaders. It is exploratory from a measurement standpoint because comprehensive assessment leadership instruments have yet to be formally developed and empirically validated. Thus, the instrument used in this study, an Internet-based survey, served as the foundation for future research in this area. In all, the outcomes were intended to both extend the knowledge base of assessment leadership among school and classroom leaders and provide preliminary outcomes for a measurement tool in assessment leadership.

Research Questions and Methods

As explained, the primary goal of this study was to explore assessment leadership by measuring key factors that have been suggested to influence assessment leadership practices, defined in this study as assessment learning experiences, beliefs, and

knowledge. A secondary goal was to examine the relationships among these variables when controlling for school role (school or classroom level) and school type (elementary or secondary). Three research questions were developed to accomplish these goals. A non-experimental, correlational research study was designed to answer the questions:

> RQ1. To what degree do assessment learning experiences, beliefs, and knowledge influence the assessment leadership practices of school and classroom leaders?
>
> RQ2. To what extent does school role (i.e., school or classroom leader) moderate the relationship between assessment learning experiences, beliefs, and knowledge and assessment leadership practices of school and classroom leaders?
>
> RQ3. To what extent does school type (i.e., elementary or secondary) moderate the relationship between assessment learning experiences, beliefs, and knowledge and assessment leadership practices of school and classroom leaders?

Target Population

The target population for this study were school and classroom leaders (designated as principals, assistant principals, or PLC facilitators) in one large school district in Florida. This district was selected as the research site because of its size and recent emphasis on standards-based reforms that included implementation of comprehensive assessment systems within a PLC framework. A comprehensive sampling approach was used at both elementary and secondary levels (McMillan & Schumacher, 2010). All individuals in the selected school district with the titles of principal, assistant principal, or PLC facilitator were included in the study in order to maximize response rates and comprehensively represent the total target population.

Instrumentation

Data were collected using an Internet-based survey administered via Qualtrics. The *Assessment Leadership* (ASLS), developed by the researcher using multiple sources, consists of five sections (see Appendix B). Section 1 contained items related to the

respondent's background (e.g., school level, role, history of assessment-related professional learning experiences) as well as questions about assessment learning experiences including postsecondary courses taken during preservice preparation and number of professional development sessions taken during inservice activities. Section 2 included 14 items adapted from a principal self-reflection tool as a perceived measure of *assessment beliefs* (Noonan & Renihan, 2006). Section 3 included 18 items, also adapted from the principal self-reflection tool, as a reported measure of *assessment leadership practices*. Section 4 comprised a 35-item inventory that evaluated assessment knowledge based on seven *assessment literacy competencies* (Mertler & Campbell, 2005). Section 5 collected information on respondent demographics and educational background.

Prior to administration, the assessment leadership survey was reviewed by two university faculty members in educational leadership and measurement for item construction and content. The items were derived from the principal reflection tool, organized by section and randomized within each section. Items that contained the same question stem were grouped and then randomized within groupings. This design was intended to reduce duplication in the item presentation, improve ease and clarity of administration, and reduce item response sets (i.e., respondents not attending to items). The items were then field tested with nine aspiring leaders to examine timing and address any mobile or computer administration issues. The field test was completed within projected time of approximately 30 to 60 minutes. Adjustments were made to item presentation to improve readability on both mobile devices and computer screens.

Data Collection

Once the survey was finalized, all eligible school and classroom leaders in the host Florida district were sent an invitation to complete the Internet-based survey via their district electronic mail addresses. The survey was administered in the spring semester following the fourth full year of the school district implementing the PLC model. The survey was left open for six weeks, and during that time an electronic mail prompt was sent each week reminding qualified personnel to complete the survey or to finish the survey if not yet completed. At the close of the survey, all complete and incomplete data were captured for analysis.

Data Analyses

Several quantitative analyses were planned and conducted to address the research questions in this study. Figure 1.1 illustrates the progression of analyses followed. Foremost, descriptives were collected for the final sample across key variables. Next, a series of factor and classical item analyses were conducted to identify the best model fit for each variable. Last, multiple regression and hierarchal regression formulas were computed to measure the influence of independent variables on the dependent variable.

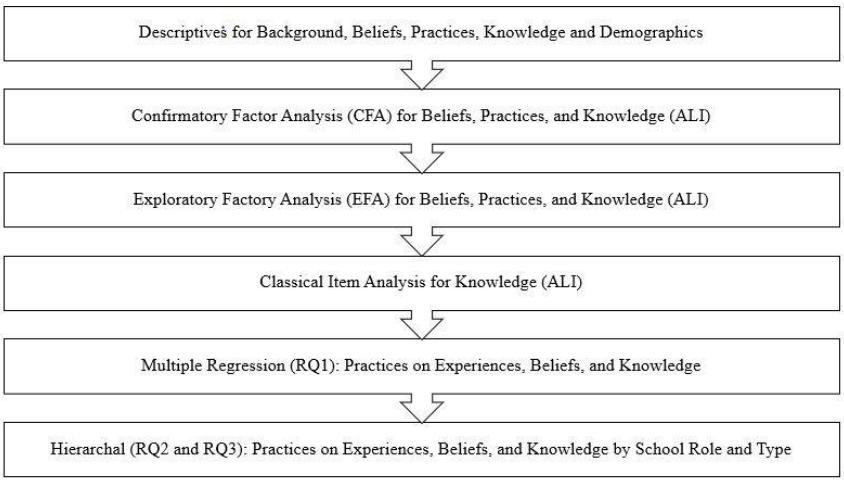

Figure 1.1. Data analyses conducted to address the research questions.

In preparation for analyses, it was noted that there were significant missing data at both survey instrument and item levels. In all, 92% of the survey variables and 79% of the respondents had at least one missing value. Contemporary methods for handling missing data were followed (Acock, 2005; Brick & Kalton, 1996; Brown, 2015; Enders, 2010; Dong & Peng, 2013; Johnson & Young, 2011; Liu & De, 2015; Manly & Wells, 2015; Pampaka, Hutcheson, & Williams, 2015; Schafer & Graham, 2002; van Buuran, 2007). First, the patterns of missing data were examined. Both general and monotone were present with successive amounts of missing data as the survey progressed, suggesting respondents did not complete the entire survey. The missing data were determined as not completely at random, but rather, at random and seemingly associated with variables within the survey. Two approaches were taken to handle the missing data. First, total nonresponses (i.e., cases with no data) and limited partial nonresponses (i.e., cases with very limited data) were removed from the original sample (Brick & Kalton, 1996). Second, 100 multiple imputed datasets were created using the remaining 284 cases from the final sample and then pooled for descriptive and inferential analyses.

Next, descriptive statistics were calculated for background items and demographic information as well as for the means and standard deviations of the survey items. Pooled means were compared using independent samples t-tests across items for assessment learning experiences, beliefs, practices, and knowledges to examine trended responses. Then, factor and item analyses were conducted using the multiple imputed datasets to examine the factor structures of the beliefs, practices, and knowledge measures. Confirmatory factor analysis (CFA) with maximum likelihood parameter (MLR) estimators was conducted in *Mplus Version 8* using the 100 imputed datasets. Four fit

indices were reviewed to evaluate the stability of the measure (Brown, 2015; Jackson, Gillaspy, & Purc-Stephenson, 2009). Due to poor fit, Exploratory Factor Analysis (EFA) using Principal Axis Factoring (PAF) with Direct Oblimin rotation and eigenvalues > 1.0 in IBM®SSPS®. The imputations were analyzed individually and outcomes averaged across imputations for trends. Outputs were reviewed for assumptions, number of factors, factor loadings, and reliability. Then, EFA was rerun based on the observed trended factor structures after item deletion. Last, classical item analysis was employed for the assessment knowledge (ALI) measure to inform item stability and reliability.

Once the factor structures were finalized, multiple regression analyses were performed to answer the research questions, specifically to explore the influence of the assessment learning experiences, beliefs, and knowledge variables on assessment leadership practices (Enders, 2010). Total scores of the items identified in the factor and item analyses were calculated and used in the final model. Last, a hierarchal regression analysis was conducted to determine whether school role and school type moderated the influence of assessment variables on practices. To assist with providing context for interpreting the results, a document review was conducted of the target district policies and implementations pertaining to assessments and professional learning communities.

Limitations

Several limitations were evident in this study. Foremost, fewer respondents than anticipated completed the entire survey. The smaller sample size impacted the strength of analyses in addition to the generalizability of the results to the population beyond the target school district. This may have been due to several factors such as timing of the survey administration, length of survey, and respondent motivation to complete the

survey. In addition to the small sample size, there were significant missing data, which further reduced the sample size. Multiple imputations were conducted to correct for the missing data; however, this approach presented significant limitations to the interpretation of both the factor structures and relationships among variables. Last, the survey was constructed from the literature, but many of the items were considered exploratory in nature and likely need revision. Moreover, the nature of the study required participants to self-report on variables, which may overestimate or underestimate the true values. The outcomes of this study, in light of the limitations, offer recommendations for future quantitative and qualitative research in this area.

Key Terms

Key educational terms with definitions for this study are listed in Table 1.1. They reflect concepts of assessment leadership founded in research and practices as well as operational descriptions of school and classroom leaders for purposes of this study.

Table 1.1

Definitions of Key Educational Terms Used in This Study

Term	Definition
Accountability policies	State and district policies established in response to accountability legislation.
Assessment beliefs	Attitudes that assessment practices are essential components of school-improvement efforts.
Assessment learning experiences	The number of assessment-related professional learning opportunities during preservice and inservice education. In this study, assessment learning experiences are defined as the number of preservice postsecondary courses plus the number of inservice professional development sessions.
Assessment knowledge	Competencies required to be assessment literate.

Table 1.1 (continued)

Assessment leadership practices	Reported assessment activities employed within an instructional leadership framework such as establishing an assessment vision with clear expectations for student assessment systems; leading data discussions; fostering assessment literacy among teachers through ongoing, collaborative learning experiences, and self-reflecting on assessment practices.
Assessment literacy	The ability to organize, analyze, and assimilate data for evaluating and adjusting curriculum and instructional practices to meet student needs.
Assessment reform	Shifts in school and classroom leader's behaviors to use of a variety of assessment tools to inform decisions about standards-based instruction.
Assessment vision	Clear expectations for administrators and teachers to collect and use various forms of assessment data to inform instructional decisions.
Classroom assessment practices	Assessment activities implemented at the classroom such as identifying student-learning targets; creating formative and summative assessments matched to targets; collecting and analyzing student data at designated intervals; and making adjustments based on student data.
Classroom leaders	Teachers who are designated as PLC facilitators and have received specific professional learning in and are allocated time and resources for leading PLC activities among teacher teams.
Contextual conditions	State and local policies, district and school-improvement plans, professional development opportunities, professional growth systems, assessment data systems, and other instructional resources used to implement school reform.
Data discussions	School and classroom leaders engaging teacher teams in PLCs to build assessment literacy, develop assessment tools linked to learning targets, and analyze and use student data to inform curriculum and instructional decisions.
Data-informed decision-making	The use of multiple sources of student assessment data to analyze student, classroom, and school trends and make strategic decisions about curriculum and instruction with the shared goal of improving student achievement.

Table 1.1 (continued)

Individual factors	Demographics (i.e., gender, age, years as educator), experiences (i.e., preservice and inservice education in assessment) and role (i.e., principal, assistant principal, PLC facilitator).
Professional Learning Community (PLC)	Regularly scheduled forums where educators engaged in professional development centered on specific problems of practice or reform intiatives.
School factors	School type (i.e., elementary or secondary) and school leader role (i.e., principal, assistant principal, PLC facilitator).
Self-reflection	School and classroom leaders dedicating time to reflecting on their own assessment literacy knowledge, skills, and leadership practices.
School leaders	School-level administrators classified as principals and assistant principals.
School type	School level, either elementary (Pre-Kindergarten through Grade 5) or secondary (Grade 6-Grade 12 plus adult education).
Summative assessments	Assessments that occur after instruction with the purpose of measuring mastery of content. May include unit tests, final exams, end of course assessments, and statewide assessments.
Student achievement	Proficiency on grade level standards or above-based on statewide assessments.
Teacher development	School leaders develop teachers as assessment literate leaders who engage in data discussions and use data to make instructional decisions.

Key statistical terms with definitions for this study are listed in Table 1.2. They reflect statistical concepts and procedures followed to execute the research design.

Table 1.2

Definitions of Key Statistical Terms Used in This Study

Term	Definition
Confirmatory Factor Analysis (CFA)	Statistical procedure used to verify the factor structure of survey items using various fit indices.
Classical Item Analysis	Statistical procedure used to evaluate individual items to determine the reliability of the items.
Exploratory Factor Analysis (EFA)	Statistical procedure used to uncover the factor structure of survey items using factor loadings.
Fully Conditional Specification (FCS)	MI method that assumes complex relationships among variables and creates imputations based on multiple iterative sets of regression equations.

Table 1.2 (continued)

General patterns	Missing data patterns throughout the survey that appear random and not corrected with variables.
Hierarchal Multiple Regression (MR)	Statistical procedure to examine if specific variables of interest explain a significant amount of variance on the dependent variable.
Item nonresponse	Missing data category that occurs when respondents do not respond to a few of the items.
Latent patterns	Missing data pattern that reflects missing latent values for the entire population sample.
Missing at Random (MAR)	Missing data mechanism that assumes data are independent of the missing values, but may be dependent on other observed values.
Missing Completely at Random (MCAR)	Missing data mechanism that assumes data are independent of both the observed and missing values with an equal likelihood of missingness.
Missing Not at Random (MNAR)	Missing data mechanism that assumes data are dependent on both missing and observed values.
Missing Value Analysis (MVA)	Statistical analysis of missing data in a dataset.
Monotone patterns	Missing data patterns that occur when a respondent stops responding to the survey.
Multiple Imputation (MI)	Procedure for handling missing data that creates multiple complete datasets by making statistical inferences about the missing data.
Multiple Regression (MR)	Statistical procedure used to examine relationships among variables in a study.
Multivariate normal imputation or Joint Modeling (FM)	MI method that assumes joint multivariate distribution of all variables and creates imputations based on pre-specified distributions.
Noncoverage	Missing data category that occurs when a faction of the target population is not represented in the sampling population and thus cannot respond.
Partial item nonresponse	Missing data category that occurs when respondents stop responding to the survey.
Planned patterns	Missing data patterns that occur when researchers distribute portions of the survey to decrease the number of items each respondent must complete.
Pooling	Aggregation of multiple imputations for analysis.
Principled methods	Statistical approaches to handling missing data robust to reducing bias compared to traditional methods of listwise or pairwise deletion methods.
Total nonresponse	Missing data category that occurs when respondents do not respond to any of the items.
Unit nonresponse patterns	Missing data patterns for portions of a survey representing more than one variable.
Univariate patterns	Missing data pattern isolated to a single variable.

Summary

School accountability policies continue to increase demands on states to ensure school-reform efforts accelerate student-learning outcomes. In an era of standards-based and comprehensive assessment systems, data-informed decision-making processes are central components of effective instructional models. As such, principals as instructional leaders serve a vital role in assisting teachers and teacher teams with developing and implementing effective student assessment and data use practices. However, gaps exist in research and practice pertaining to factors or conditions that contribute to effective assessment leadership. Given the significance of data-informed decision-making in policy and practice, knowledge about and skills in assessment leadership are critical to enhancing student-learning outcomes and achieving school-improvement goals. Therefore, continued study in this area is necessary to expand the knowledge base and inform professional development needs to improve assessment leadership practices in school and classroom leaders.

Chapter 2 presents an overview of theoretical models for instructional leadership and evidences of assessment reform in an era of school accountability policies and practices. Research on shifts in classroom and leadership practices in the areas of assessment and data-informed decision-making practices are presented as the foundation for conceptual framework and research on assessment leadership. Chapter 3 provides specific details about the research design and data collection and analysis techniques. Chapter 4 presents the results of the data analysis within the context of the school district's initiatives in the areas of assessments and PLCs. Chapter 5 summarizes key findings and makes connections to future educational implications.

CHAPTER 2

LITERATURE REVIEW

The enactment of NCLB, followed by the reauthorization of IDEA, elevated expectations for school accountability (USDE, 2002, 2004) with large-scale, high-stakes assessments and sanctions and rewards linked to student achievement for all students (Darling-Hammond, 2004). In the years following passage of these legislations, most states adopted or adapted heightened academic standards, extending from contemporary movements in Common Core State Standards (CCSSO, 2010) and 21st century learning skills for career and college readiness (NGA, 2010). In addition, many states revised high-quality teaching criteria to incorporate student-learning outcomes as critical components of teacher evaluations (Clark, 2011; Waters, Marzano, & McNulty, 2003)

In the wake of school accountability policies, standards-based reform, coupled with fiscal incentives such as Race to the Top grants (USDE, 2009), increased pressures for districts to explore new instructional models and educational practices at classroom and school levels. Most recently, districts experienced continued momentum in school accountability with the latest iteration of federal legislation, *Every Student Success Act* (ESSA). ESSA retains the focus on student achievement in P-12 education while expanding emphases on college and career preparedness and innovation (USDE, 2015). In the first decades of the 21st century, public education policies have led to critical shifts in educational practices from teaching content alone to monitoring student-learning outcomes toward achieving high academic standards for postsecondary goals.

Shifts in educational policies and practices have translated to myriad implications for schools with a spotlight on adopting standards-based assessment tools that monitor

student-learning outcomes toward grade-level standards (Hattie, 2009; Noonan & Renihan, 2006; Stiggins & Duke, 2008). Consequently, more so now than in prior decades, schools have engaged in regularly collecting and reflecting on student data to assist with instructional decisions at classroom, school, and district levels. These assessment-related activities increasingly have influenced the roles and responsibilities of principals beyond the traditional administrator to include creating and using data systems, leading data discussions pertaining to instructional practices, conducting personnel evaluations linked to student-learning outcomes, and using standards-based data to inform school-improvement plans (Means, Padilla, & Gallagher, 2010; Noonan & Renihan, 2008; Young & Kim, 2010). As such, it is essential to understand the conditions necessary for principals to successfully navigate their evolving role to meet the heightening expectations of educational reform, specifically in the areas of student assessment and data use (Sanders & Kearney, 2008; Stiggins & Duke, 2008).

This chapter explores models of instructional leadership that serve as the theoretical bases for study in assessment leadership. The current knowledge base in standards-based assessment systems as well as in principal and teacher leader behaviors are also reviewed. These components reflect shifts in assessment practices from traditional approaches to standard-based and embedded professional learning communities' approaches as functions of evolutions in school accountability policies and educational reforms. In all, these research and practice bases underscore the conceptual framework for study in assessment leadership.

Theoretical Framework

While researchers have explored the construct of instructional leadership for decades (Hallinger & Heck, 2010; Lashway, 2003; Thompson, 2012), heightened expectations for student achievement required by school accountability movements, followed by innovations in educational practices, have refocused principals' work on leadership activities that promote learning and teaching (Darling-Hammond, 2004; Fullan, 2001; Green, 2010; Noonan & Renihan, 2006). As such, principals increasingly are expected to envision, facilitate, support, and monitor instructional practices as central components of day-to-day practices (Hallinger & Heck, 2010; Valentine & Prater, 2011). These shifts in principal practices also have prompted movement toward more collaborative leadership models that distribute leadership roles and responsibilities by engaging all teaching professionals in shared decision making around student-learning outcomes (Green 2010; Hattie 2009; Leithwood & Louis, 2012; Sanders & Kearney, 2008). At the center of collaborative instructional leadership practices are assessments and data use, which provide the foundation for research on assessment leadership (Noonan & Renihan, 2006, 2010; Stiggins & Duke, 2008).

Instructional Leadership

The concept of instructional leadership originated in the 1980s from research on effective school models. Traditionally, instructional leadership was conceptualized as principals' knowledge about and experience in curriculum and instruction, such as spending time in classrooms observing, providing feedback, and modeling effective strategies (Hallinger, 2001). Over the last three decades, researchers have explored various definitions, standards, and frameworks for instructional leadership (Cotton, 2003;

Hallinger, 2001; Marzano, Waters, & McNulty, 2005). As such, instructional leadership has evolved to include both the narrow scope of principals conducting observations and providing feedback in the classroom to the broad scope of principals and teachers establishing a vision for creating a culture of educational practices centered on learning and teaching in the school (Leithwood & Louis, 2012).

Instructional leadership entails principals intentionally developing effective learning environments that promote high expectations and sound instructional approaches (Fullan, 2001; Robinson et al., 2008). Hallinger (2001) delineated a three-component model that reflects the core aspects of instructional leadership: define the school's mission, manage the instructional program, and promote a positive school culture. Within this model, principal activities involve (a) developing strategic school-improvement plans that incorporate instructional goals and data monitoring systems, (b) possessing high levels of instructional knowledge, (c) guiding and evaluating instructional programs, (d) allocating resources toward development of curriculum and assessments, and (e) serving as a change agent that encourages members to engage in ongoing professional learning opportunities to build skillsets in instructional practices. Such activities represent an expansion of the principal role beyond classroom observations alone to include fostering school environments of learning and teaching.

Like Hallinger (2000), Horng and Loeb (2010) also advocate for an expanded definition of instructional leadership, arguing that principals who lead primarily by working directly with teachers are not as effective as principals who lead by fostering a school environment that facilitates and supports teacher development in curriculum, assessment, and instruction. Within such school environments, principal roles and

responsibilities encompass developing curriculum and assessment systems, providing professional learning opportunities, and executing teacher effectiveness systems (Jenkins, 2009; Leithwood, Harris, & Hopkins, 2008). In this model, principals serve a central role in creating a school culture that is jointly focused on the professional development of teachers and the academic success of all students. This approach to instructional leadership requires a sizable shift in leader practices where principals dedicate time and resources to developing organizational infrastructures that support learning and teaching.

Although a universally accepted definition of instructional leadership is not comprehensively established, common themes of effective practices have emerged, following which, many states have established standards for principal practices (Thompson, 2012). Of note, the Interstate Principal Licensure Consortium (ISLLC) led this effort by identifying six core standards that promote student success in 21st century learning environments (Saunders & Kearney, 2008). The standards center on core concepts of instructional leadership that include (a) establishing a shared vision, (b) creating a culture of student and staff learning, (c) establishing and maintaining safe and effective learning environments, (d) collaborating with the broader community, (e) acting ethically, and (f) advocating for educational excellence within broader contexts. These standards encompass shared expectations for leadership practices with the goal of sustaining positive effects on school outcomes. Thus, they have been adopted or adapted by most states and organizations as the basis for principal preparation programs (Hallinger, 2001, 2011; Young, Crow, Murphy, & Ogawa, 2009).

Transformational Leadership

Like instructional leadership, the concept of transformational leadership has been well established in educational research, initially introduced in the late 1970s. Transformational leadership emerged from Burns' (1978) research on leadership practices that promote change and momentum through shared vision, mutual benefit, and inspiration. This model of leadership involves leaders and followers developing a partnership that fosters trust and promotes shared roles and distributed responsibilities (Green, 2010). Leaders engage followers using proactive strategies that activate the collective interests of the group through influence, stimulation, and motivational approaches (Hoy & Miskel, 2008). Such an approach connects educators on both intellectual and emotional levels and works to both establish and maintain motivation for principals and teachers to participate in productive and effective work.

Burns' (1978) initial work was continued in later decades by Rost (1991) to incorporate collaborative components that involve followers sharing in responsibilities, contributing to the common goals of the organization. Rost conceptualized leadership as "an influence relationship between leaders and followers who intend real changes that reflect their mutual purposes" (p.102). These practices are differentiated from management in that both leaders and followers are involved in a relationship based on influence with the intent on making organizational changes based on mutual goals. Through a shared commitment to leadership, leaders and followers accomplish transformations in organizational practices that impact all professionals.

The cornerstone of transformational leadership is the development of a clear vision that is shared by all stakeholders (Green, 2010). Leaders persuade followers to

accept their vision and combine efforts to accomplish mutual purposes (Rost, 1991). Inherent within the shared vision is the implication that transformational leaders impart leadership by distributing power and authority to the followers (Green, 2010; Rost, 1991). Given this definition, principals as transformational leaders must ascertain the strengths and needs of the organization and then position teachers and staff members to perform leadership functions that maximize their potential through utilizing their interests and abilities. A key component to distributing leadership within this model is for leaders to identify and match followers that will not only delegate the work but also translate it into improved student-learning outcomes (Collins, 2001). This form of leadership communicates trust and engages all stakeholders as responsible and accountable contributors to the common vision and goals of the school.

More contemporary perspectives on transformational leadership posit that such approaches to instructional leadership are even more essential in 21st-century educational environments given the need to generate and support the instructional changes drive by school reform (Green, 2010; Noonan & Renihan, Stiggins & Duke, 2008). Principals must consider strategies such as collaborative practices and ongoing professional development of teachers as additional components of their work requirements (Leithwood, 1992). These approaches extend beyond instructional leadership alone to engage professionals actively in activities that share a common vision and serve to innovate and accelerate learning not only for students but also for teaching professionals.

Collective Leadership

Different from instructional and transformational leadership models, collective leadership is an emerging framework that more narrowly focuses on the teaming

processes necessary to distribute and operationalize work in school environments (Hallinger & Heck, 2010; Lewis, Leithwood, Wahlstrom, & Anderson, 2010). However, similar to transformational leadership, collective leadership assumes leadership is bi-directional between principals and teachers and requires the involvement of leader representatives from all levels of influence as contributors to instructional and school-climate decisions and outcomes (Hattie, 2009; Leithwood & Louis, 2012). Consequently, in this model, leadership roles and responsibilities shift away from the individual and capitalize on the collective capacity and mutual accountability of the group to leverage the work. This leadership approach also has been reported by principals as effective in facilitating assessment practices in schools (Rehihan & Noonan, 2012).

Collective leadership encompasses instructional activities, but more importantly emphasizes how principals and teachers engage in collaborative, data-informed decision-making to impact learning (Leithwood & Louis, 2012) to create a culture of inquiry that is grounded in research-recommended practices (Boudett et al., 2010; Green, 2010). Similar to work within communities of practices (Wenger, 1998), collective leadership hinges on the capacity of group members to engage in learning and embed professional knowledge into educational practices and decisions (Collinson & Cook, 2007). In school environments, this approach is best characterized as PLCs: teacher teams guided by the shared vision of its members (Danielson, 2009; Dufour & Eaker, 1998). Like in models of transformational leadership, PLCs functioning as communities of practice (Wenger, 1998) require commitment, trust, and contribution by all participants to generate productivity and yield effective outcomes for the organization.

While collective leadership has been suggested as a viable approach to

instructional leadership, differences are evident across school type and role. For example, Gedik and Bellibas (2015) conducted a comprehensive study of distributed leadership among teachers and administrators in elementary and secondary schools nationwide. They measured critical leadership practices in (a) maintaining a focus on learning, (b) monitoring learning and teaching, (c) building nested learning communities, (d) acquiring and allocating resources, and (e) maintaining safe and effective learning environments. The results suggested that elementary school staff tended to be more engaged in instructional leadership practices than secondary school staff. Similar results were observed for administrators compared to teachers. Differences between levels and roles were particularly evident for monitoring learning and teaching, including conducting formative and summative evaluations, at elementary schools compared to secondary schools. These differences may be explained by variables such as time spent on leadership activities, resource allocations, school infrastructures, and staff perceptions about leadership roles and responsibilities (Gedik & Bellibas, 2015).

Although collective leadership is newly conceptualized in research and practice, it is positioned as a critical element of school leadership that encompasses a culture of learning, teaching, and data-informed decision making while also meeting the multiple demands of managing a school (Gedik & Bellibas, 2015). The cumulative effects of collective-leadership practices support instructional engagement and maximize student achievement outcomes (Hallinger & Heck, 2010). Further, the deliberate and systematic involvement of teachers in job-embedded professional learning advances the knowledge and skills of teachers, with the goal of retaining them (Collinson & Cook, 2007).

Case for Assessment Leadership

It is evident that evolutions in accountability policies have influenced assessment reform in states, districts, and schools. At the forefront of implementing these policies are principals and classroom teachers making intentional strides to become more data oriented by (a) adopting student-assessment systems that contain varied standards-based assessment tools collected and (b) using those results to inform instructional purposes throughout the school year (Coburn & Talbert, 2006; Ingram et al., 2004; Kerr at al., 2006; Means et al., 2009). Increasingly, principals and classroom teachers have recognized the value of formative assessments and employing a variety of techniques to measure student learning outcomes compared to curriculum standards (Black et al., 2004; Loeb et al., 2008; Volante & Beckett, 2011), which has led to shifts in teaching and leadership practices, specifically in the area of assessment.

Shifts in Classroom Assessment Practices

Historically, teachers have engaged in a variety of assessment techniques, from informal observations to formal tests (Young & Kim, 2010). Formal tests have typically been used to assign grades, and day-to-day student interactions have been recorded as a reflection of learning (Frey & Schmitt, 2010; Goertz, Olah, & Riggan, 2009). Such assessment tendencies have resulted in an overemphasis on summative tests to gauge learning, which reflects only if students have mastered the content after instruction has occurred (Frey & Schmitt, 2010; Volante, 2010). Summatives alone allow few opportunities for teachers to use data to evaluate student learning *during* instruction (i.e., before assigning a grade and proceeding to new content in the curriculum). In recent years, teachers have acknowledged that standards-based classroom assessments are

important to student success and are needed more often than once or twice a year (Guskey, 2003; Marzano, 2006; Stiggins, 2002) beyond the scope of traditional uses for summative assessments (Loeb et al., 2008). As a result, classroom assessment practices have evolved to incorporate a variety of assessment tools and techniques used to inform instructional decisions throughout the school year (Lee & Wiliam, 2005; Marzano, 2006; Suurtamm et al., 2010; Young & Kim, 2010). These practices include making strategic adjustments to classroom assessments to align with new content standards and district and state assessments in response to student-learning outcomes (Loeb et al., 2008).

The purpose of classroom assessments also has progressed from teachers recording results in their gradebooks to their using the results to adjust instruction in response to student-learning needs, including students monitoring their own learning (Black, Harrison, Lee, Marshall, & Wiliam, 2004; Stiggins, 2005). While traditional assessment practices have encompassed some form of formative data collection, it has primarily involved informal, anecdotal observations (Goertz et al., 2009) and has not been sufficient enough to inform instructional decisions that yield improved student learning-outcomes (Stiggins, 2002). Formative assessment practices that are directly tied to standards-based learning targets are the most effective modes of evaluating instruction to make adjustments prior to summative assessments (e.g., unit tests, semester exams) (Black et al., 2004). Shifts in the purpose of assessments requires extensions beyond cursory classroom observations of learning toward a more standardized approach based on learning criteria linked to the curriculum standards (Shavelson et al., 2008).

In addition, the number and nature of classroom assessments have advanced in frequency and type to incorporate more regular intervals of data collection at multiple

levels (e.g., student, classroom, school, district) (Stiggins, 2005). 2007; Preuss, 2007; Shen et al., 2010). These assessments range in purpose from ones that occur in the moment to planned-for-interactions and embedded-in-the-curriculum assessments that are more formal in nature and intentionally designed prior to instruction (Shavelson et al., 2008). They include a variety of assessment techniques such as questioning, self-assessments, and formative use of summative measures such as grades and final exams to examine curricular trends (Black et al., 2004; Stiggins, 2002; Suurtamm et al., 2010). Given the diverse classroom assessment data collection strategies and tools, formative assessments are not so much defined by the exact assessment technique but rather by how teachers use the results to make critical instructional decisions (Wiliam & Leahy, 2006).

Shifts in Leader Assessment Practices

Traditionally, leaders have relied on assessment data to establish annual goals, guide professional developmental activities, and inform district- and school-improvement plans (Shen et al., 2010). However, most often, these data have been derived from state assessments that occurred in previous years with minimal influence on current-year learning targets. In recent years, principals have reported viewing ongoing assessment data as more useful in their work, including integrating data more frequently and extensively into school-improvement planning as well as providing teachers with more resources and support for accessing and using student assessments to guide instruction (Kerr et al., 2006; King & Amon, 2008; Marsh, Pane & Hamilton, 2006; Supovitz & Klein, 2003; Young, 2006). Such shifts in school leadership practices are consistent with shifts observed in classroom teachers in the area of standards-based formative assessments. While time and access to data present challenges, principals praise

assessment-reform efforts more positively than negatively as prompting them to be more effective instructional principal and teacher leaders (Prytula, Noonan, & Hellsten, 2013).

Changes in assessments have not only affected school leadership practices but also influenced interactions among principals and teachers (Coburn & Talbert, 2006; Kerr et al., 2006; Loeb, Knapp, & Elfers, 2008; Shen et al., 2010). More often, principals are using data to make strategic changes in resource allocations as well as the grouping and regrouping of students in response to the data (Halverson et al., 2007). In addition, teachers and principals spend more time working collaboratively in PLCs and data conferences to make data-informed decisions about needed instructional modifications (Suurtamm, Koch, & Arden, 2010). They are engaged in ongoing assessment and data-use activities such as (a) data acquisition, analysis, and reflection; (b) instructional design and curricular alignment; (c) use of both formative and summative assessments; (d) collaborative problem-solving discussions and action planning; and (e) targeted professional development in these areas with regular communication and corrective feedback (Arter, Stiffins, Duke, & Sagor, 1993; Blink, 2007; Coburn & Talbert, 2006; Halverson et al., 2007). Increasingly, principals and classroom teachers alike are dedicating time and resources to using formative, standards-based data to guide instructional decisions at student, classroom, and school levels.

Barriers to Effective Assessment Practices

Student assessment systems that incorporate formative practices provide a viable avenue for principals and teachers to collect and use assessment data to inform instruction practices with the goal of accelerating student-learning outcomes (Black & Wiliam, 1998; Stiggins, 2002; Stiggins, 2005). Unfortunately, several barriers inhibit

successful implementation of administering, scoring, and analyzing data for instructional purposes (Ingram et al., 2004; Shen et al., 2010). Key obstacles include available time and access to relevant data for instructional use, differences in educational beliefs and philosophies among collaborators, and lack of professional development in appropriate assessment tools and data use (Burke & Wang, 2010; Coburn & Talbert, 2006; Deenen & Brown, 2016; Gallagher et al, 2008; Kerr et al., 2006).

At the forefront is feasibility of student assessment collection and organization for use. Formative assessments of student learning typically are conducted using paper and pencil, which requires more time to score and may not lend themselves to quick analysis for instructional decisions (Hall & Hewitt-Gervais, 2000). Even when the results are immediately available, teachers often lack time or expertise to make adequate sense of the data for improving or altering their instruction. As a result, teachers tend to default to summative assessments, such as quizzes and tests that occur after instruction has been delivered and for grading purposes (Penuel, Tatar, & Roshelle, 2004; Roschelle, Penuel, Schechtmann & Tatar, 2005). Such approaches are not as effective as establishing a systematic approach to formative assessment and analysis.

Inadequacies in student assessment data systems also contribute to problems related to accessing and using data for instructional purposes (Gallagher et al., 2008; Wayman, 2005, Ulmer, 2002). Typical data systems serve as management tools for housing data such as student demographics, assessment participation rates, and test scores. Teachers reported that such data have limited utility and relevance in assisting them in making instructional decisions (Means et al., 2009) due to the time and multiple

steps required to access usable information for classroom use. Thus, data within such systems have little impact on teachers' decision making about needed changes.

In addition to limited access to efficient data sources, many teachers struggle with examining instruction using a variety of student assessment sources and decision-making protocols (Boudett et al., 2010). They tend to proceed to action planning before they have effectively analyzed the data and developed a clear understanding of the problem or issue. Many schools have instituted PLCs, common planning time, data conferences, and other forums to assist teachers in evaluating assessment results and sharing strategies for instructional improvement in collaboration with other professionals (Goertz et al., 2009; Means et al., 2009). However, even these approaches do not always provide sufficient time and resources for teachers to use assessment results when reflecting on and planning for instruction. Without sufficient experiences, beliefs, and knowledge, teachers tend to default to traditional practices of solely focusing on delivering the content and relying on informal classroom observations to monitor student progress (Young & Kim, 2010).

Key Components of Assessment Leadership

The convergence of research on data-informed decision-making processes with an emphasis on formative assessment systems (Black & Wiliam, 1998; Mertler, 2005; Stiggins, 2001) underscores the importance of schools adopting key instructional practices such as (a) accessing multiple sources of student assessment data, (b) facilitating reflective data discussions, and (c) using data to inform decisions at multiple levels. Moreover, it also has been established that principals are essential to instituting effective instructional practices in their schools (Gedik & Bellibas, 2015; Hallinger & Heck, 2010; Lashway, 2003; Louis et al., 2010; Thompson, 2012). Thus, it is

hypothesized that principals who adopt effective assessment leadership practices will experience more success in improving learning and teaching practices as well as increasing student achievement than by engaging in traditional leadership behaviors alone. While assessment leadership as a construct is newly conceptualized in the literature (Noonan & Renihan, 2006, 2008, 2010; Renihan & Noonan, 2012; Popham, 2009; Stiggins & Duke, 2008), key elements of instructional, transformational, and collective leadership converge to frame the theoretical basis for this construct. This definition of assessment leadership is proposed: s*chool leaders who establish inquiry-based professional learning environments that promote assessment literacy and employ effective assessment practices with the goal of improving student-learning outcomes*. The following sections describe each of the major component of this definition.

Inquiry-based learning environments. While principals are essential to the development of effective student assessment and data use practices in schools (Boudett et al., 2010; Noonan & Renihan, 2006; Halverson et al., 2007), implementation is not accomplished by the principal alone (Knapp et al., 2006; Noonan & Renihan, 2006). Rather, assessment leadership involves building assessment literacy among teaching professionals while also creating a culture of inquiry that promotes communities of practice (Copeland, 2000; Fullan, 2001; Gedik & Bellibas, 2015; Hattie, 2009; Supovitz & Klein, 2003; Wenger, 1998). Inquiry-based learning environments require principals and teachers to adopt a common vision for effective student assessment and data use practices with distributed roles and responsibilities among all stakeholders (Bernhardt, 2004; Green, 2010; Sanders & Kearney, 2008). This includes arranging time, resources, and professional learning opportunities for teachers not only to create and refine

assessment tools and data use strategies but also to learn about and engage in professional conversations around student achievement (Danielson, 2009; Dufour & Eaker, 1998).

By dedicating resources to developing and reflecting on effective student assessment and data use practices, principals communicate to all stakeholders expertise about and investment in curriculum and instruction with guidance on how to balance school accountability, professional values, and student needs (Mintrop, 2012). The cumulative effects of shared leadership help to establish consensus and contribute to a sustainable culture of inquiry that yields positive learning-outcomes for students (Boudett et al., 2010; Leithwood & Louis, 2012) as well as generate leadership opportunities for teachers through the diffusion of evidence-based practices (Danielson, 2009) in safe and supportive learning settings (Dufresne & McKenzie, 2009).

Assessment literacy. The cornerstone of assessment leadership and effective student assessment and data use practices is assessment literacy (Earl & Fullan, 2003; Noonan & Renihan, 2006, 2010; Popham, 2010; Stiggins & Duke, 2008), which is the ability to organize, analyze, and assimilate data for the purpose of evaluating and adjusting instructional practices to address student-learning needs (Fullan, 2001). It requires the capacity to examine student data, develop action plans based on the data, and engage in discussions about data use. Essential skills of assessment literacy include understanding the purposes of assessments and their instructional reliability and validity, addressing personal beliefs and biases about assessments, constructing good formative assessments, aligning assessments with curriculum standards, evaluating and scoring student work, and using assessments to inform instruction (Popham, 2010; Webb, 2002).

Assessment practices. A review of the literature about assessment leadership suggests five key leadership practices: (a) establishing a vision for data use, (b) setting clear and appropriate learning targets aligned to content standards, (c) using assessment data to evaluate and adjust instructional programs matched student needs, (d) developing assessment competencies among teachers through collaborative learning experiences, and (e) engaging in ongoing self-reflection in assessments (Halverson, Grigg, Prichett, & Thomas, 2007; Loeb et al., 2008; Militello, Schweid, & Sireci, 2010; Noonan & Renihan, 2006, 2010; O'Donnell & White, 2005; Popham, 2009). Moreover, Stiggins and Duke (2008) articulate ten essential competencies of principals working as assessment leaders that reflect specific knowledge and skills they must possess and use to engage in these practices. These competencies encompass understanding diverse types of classroom assessments and their relationship to student learning, knowing how to analyze data for instructional purposes, and identifying attributes of quality assessment systems.

Challenges in Assessment Leadership

Although principals increasingly have been involved in more assessment leadership practices (Clifford & Mason, 2013; Noonan & Renihan, 2006, 2008; Prytula et al., 2013; Renihan & Noonan, 2012), they face many challenges in implementation (Stiggins & Duke, 2008; Volante & Cherubini, 2011). Foremost, principals reported feeling underprepared (Clifford & Mason, 2013; Ulmer, 2002). Specifically, they identified barriers in beliefs about data use, fit with school culture, availability of adequate assessment resources and easily accessible data warehouses, and knowledge of and training in operationalizing school-improvement plans to accelerate student outcomes

effectively (Ulmer, 2002). These factors interfere with principals' abilities to establish and support effective student assessment and data use practices in their schools.

Some districts have attempted to overcome these barriers by improving access to and efficiencies in student data systems (Means, Padilla, DeBarger, & Bakia, 2009). However, even with better systems, principals and teachers continue to identify a range of challenges including fears about data security, beliefs about data utility, general disinterest in data, inadequate knowledge and skills to perform data analysis tasks, and lack of adequate training, resources, and leadership supports (Gallagher, Means, & Padilla, 2008; Means et al., 2009; Volante & Cherubini, 2011; Young & Kim, 2010). In the face of such challenges, educators have resorted to traditional approaches to assessment data collection such as intuition, experience, and anecdotal information in place of more contemporary, validated assessment approaches (Ingram et al., 2004).

At the core of implementation issues is limited exposure to assessment courses as part of principal and teacher preparation programs and continuing educating professional development opportunities (Popham, 2010; Stiggins, 2001; Wayman, Midgley & Stringfield, 2006). While university-based preparation programs incorporate various aspects of instructional pedagogy and leadership, few programs instruct specifically in the areas of assessment literacy and assessment-specific instructional leadership practices (Bernhardt, 2004; Deneen & Brown, 2016; Stiggins, 2002; Wayman et al., 2006). Surveys of assessment literacy suggest inconsistent trends in assessment literacy levels between teachers and principals (Hameister, 2013; Matthews, 2007; Perry, 2013) and among educators who are preservice compared to those who are inservice (Alkharusi, Kazem, Al-Musawai, 2011; Beziat & Coleman, 2015; Campbell, Murphy, & Holt; 2002;

Mertler, 2005). Lack of coursework on effective student assessment and data use practices likely has contributed to variations in beliefs, knowledge, and skills among principals and teachers in elementary school and particularly in secondary school levels (Brookhart, 2001; Deneen & Brown, 2016; Henry, 2011).

With respect to assessment leadership, principals tend to demonstrate higher levels of assessment knowledge and skills for engaging in ethical assessment practices and selecting a strategy or device for data collection than for interpreting and using data to inform instructional decisions (Impara & Plake, 1995). While they report valuing formative assessment data over summative assessment data, they feel more confident in using summative data to inform decisions (Henry, 2011). Limitations in principal and teacher preparation programs and ongoing continuing education opportunities contribute to the depth of challenges districts and schools face in successfully implementing student assessment systems necessitated to realize positive learning outcomes for all students.

Conceptual Framework

As described, assessment leadership practices are evident in models of instructional, transformational, and collective leadership. These practices include the specific actions leaders must take to ensure classroom teachers effectively use data to inform their instructional decisions, and thus, serve as a foundation for assessment leadership and as a construct in research and in the schools. The conceptual framework (Maxwell, 2005) used for this study encompasses both the external expectations of accountability policies and educational reform efforts and the internal assessment leadership practices necessary to create successfully a culture of assessment practices and data-informed decision making with the shared goal of improved student achievement.

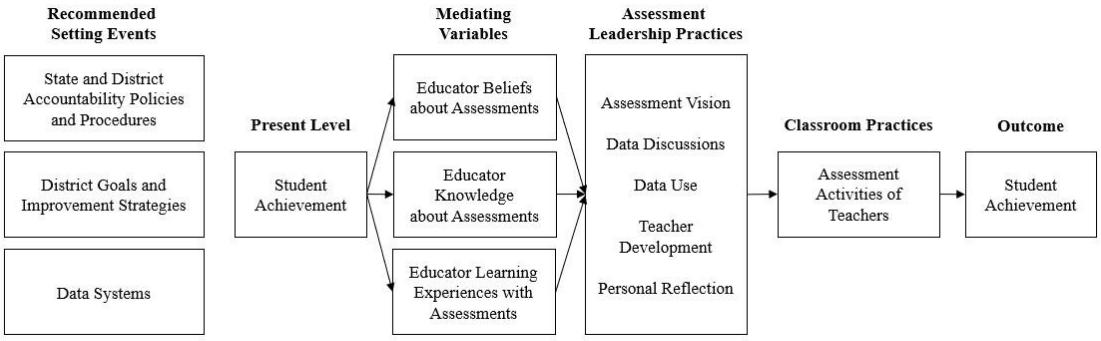

Figure 2.1. Conceptual framework for Assessment Leadership.

Figure 2.1 illustrates the assessment leadership conceptual framework established for this study. The elements contained in the framework were derived from research-based instructional leadership models, comprehensive assessment systems, principal and classroom teacher assessment practices, and state and school-district reform policies and practices in assessments. As shown, the recommended setting events comprise three components. The first reflects the state and local policies and procedures that contribute to heightened expectations for school accountability and underscore the need for school assessment reform. These factors include assessment-driven state statutes and district policies and procedures in assessment development, administration, and reporting requirements that extend from state guidelines. The second reflects the goals and strategies identified in district improvement plans designed to meet state expectations in assessment data collection and use, which sets the stage for school and classroom leaders charged with carrying out implementation in schools. The third involves development of data systems at district and school levels, which provides access to data on student assessment outcomes for making important curriculum and instruction decisions.

Assessment leadership practices were derived from research on critical components of effective leadership practices in this area (Noonan & Renihan, 2006,

2010; Stiggins & Duke, 2008). At the school level, assessment leadership involves (a) setting a vision; (b) establishing and using data systems, which includes creation of and access to formative and summative assessment tools; (c) facilitating data discussions; (d) promoting and developing of teacher leaders; and (e) self-reflecting and refining assessment competencies and practices. In models of transformational and collective leadership, assessment roles and responsibilities are distributed to classroom teachers. Thus, it is expected, in such models, that classroom leaders also would undertake principal practices, which would be influenced by their own knowledge of, beliefs about, and experiences with assessments. School-leader and classroom-leader practices leverage the capacity for classroom teachers to engage in effective assessment practices, which, in turn, increases the potential for improved learning outcomes for all students.

The current level of student achievement, as defined by student performance on statewide standards-based assessments, represents the present level and serves as the baseline of the assessment leadership framework. The outcome reflects changes in student achievement, depending on the degree to which school and classroom leaders influence the assessment practices of classroom teachers. The framework assumes that when effective assessment school and classroom leadership practices are present, classroom practices incorporate more effective assessment activities, and thus, student achievement improves. However, effective leader practices may be influenced by key mediating variables such as (a) beliefs about assessments and data use, (b) assessment knowledge (i.e., assessment literacy levels), and (c) experiences with assessment practices and data-informed decision making (i.e., postsecondary coursework and professional development workshops). These mediating variables are hypothesized to

contribute to the development of school and classroom educators as assessment leaders; thereby, influencing the likelihood of effective assessment leadership practices. The model is intended to be iterative, in that as student achievement increases as a function of assessment leadership, these assessment leadership variables are influenced positively.

Previous Studies in Assessment Leadership

Research in assessment leadership primarily has focused on measuring assessment literacy levels of teachers and principals and documenting assessment attitudes, beliefs, practices, and tools using a survey method approach (Brookhart, 2001; Deluca, LaPoint-McEwan, & Luhanga, 2016). These studies revealed few differences in assessment literacy levels among elementary and secondary administrators with varied results among teachers and principals (Davidheiser, 2013; Hameister, 2013; Matthews, 2007; Perry, 2013) and the uses of assessment data for instructional decisions (Henry, 2011). One study evidenced higher levels of assessment literacy among preservice teachers (Campbell, Murphy, & Holt; 2002), while other more recent studies consistently reported higher levels of assessment literacy among inservice teachers when compared to preservice teachers (Davidheiser, 2013; Mertler, 2005). Differences in research outcomes are due many factors including variations in educator preparation programs as well as in opportunities for ongoing professional development experiences over time.

When compared to preservice and inservice teachers, principals demonstrated lower levels of assessment literacy on the same assessment literacy inventory (Perry, 2013). Although principals tended to report valuing formative assessment data over summative assessment data, they reported feeling more competent in using summative data to inform decisions at school and district levels (Henry, 2011). Moreover, principals

demonstrated higher levels of assessment literacy for ethical practices and selecting assessment tools and devices than for interpretation and data use for informing instructional decisions (Impara & Plake, 1995), which evidenced some gaps in competencies required to lead shifts in assessment practices. More recent studies examined the reliability of assessment literacy survey items, suggesting similar trends as other studies and little variance among items when given to preservice teachers (Alkharusi et al., 2011; Beziat & Coleman, 2015). Overall, the scope of empirical research on assessment literacy surveys is limited with scant statistical evidence for assessment leadership beyond conceptual and qualitative literature reviews.

Current Study in Assessment Leadership

The need for study in assessment leadership is multi-faceted. Foremost, federal, state, and district policies continue to emphasize school accountability, requiring schools to demonstrate student achievement matched to grade-level standards on annual standards-based assessments (USDE, 2015). Evolutions in policy and practice underscore the continual need for advancements in assessment reform. Renewed demands for assessing and monitoring student-learning outcomes in policy and practice highlights assessment leadership as a relevant research topic (Noonan & Renihan, 2006; Popham, 2009; Prytula et al., 2013; Renihan & Noonan, 2012; Stiggins & Duke, 2008).

Significance. The purpose of this study is to add to the literature base on several fronts. Not only has it been shown that formative assessments and data-informed instructional practices improve student and school outcomes, but also that school leadership is essential to establishing a culture of inquiry, data-informed decision making, and professional learning for staff and students (Copeland, 2000; Popham, 2010).

Assessment literacy and other assessment leader components are critical for principals and classroom teachers, especially in an era of school accountability that call for shifts in standards-based student assessment systems (Noonan & Renihan, 2006; Popham, 2009). In addition, the collective leadership involving all school and classroom leaders serves to enhance and sustain effective educational assessment practices and school-wide improvement efforts (Green 2010; Hattie, 2009; Leithwood & Louis, 2012).

Research also has shown that educators report myriad challenges in developing and sustaining effective student assessment practices toward school improvement, including the increasing demands on principals to balance instructional leadership with traditional managerial tasks (Stiggins & Duke, 2008). Therefore, more is research needed to understand the landscape of assessment leadership practices and the needs for improvement in this area. Specifically, practice would benefit from the examination of variables that contribute to assessment leadership practices in school and classroom leaders. Evident gaps in the literature include interactions among assessment literacy and other assessment leadership variables and the measures used to examine such constructs.

Assumptions. Since the conceptual framework for assessment leadership has not yet been empirically validated, this study is exploratory in nature. Even so, several assumptions were made based on previous research in this area. It was expected that positive correlations among assessment leadership variables would be observed. Moreover, school and classroom leaders who demonstrate higher levels of educator beliefs about, knowledge of, and experiences with assessments would report heightened engagement in assessment leadership practices. In addition, due to the respective leaders' roles in the schools, as well as differing structures at the elementary and secondary levels,

differences were anticipated by school role and type. In all, the results of this study contribute to the knowledge base in assessment leadership as a construct by describing the relationships among these key variables, including how school role and type influence assessment leadership practices. The results also provide preliminary outcomes for a measurement tool in assessment leadership.

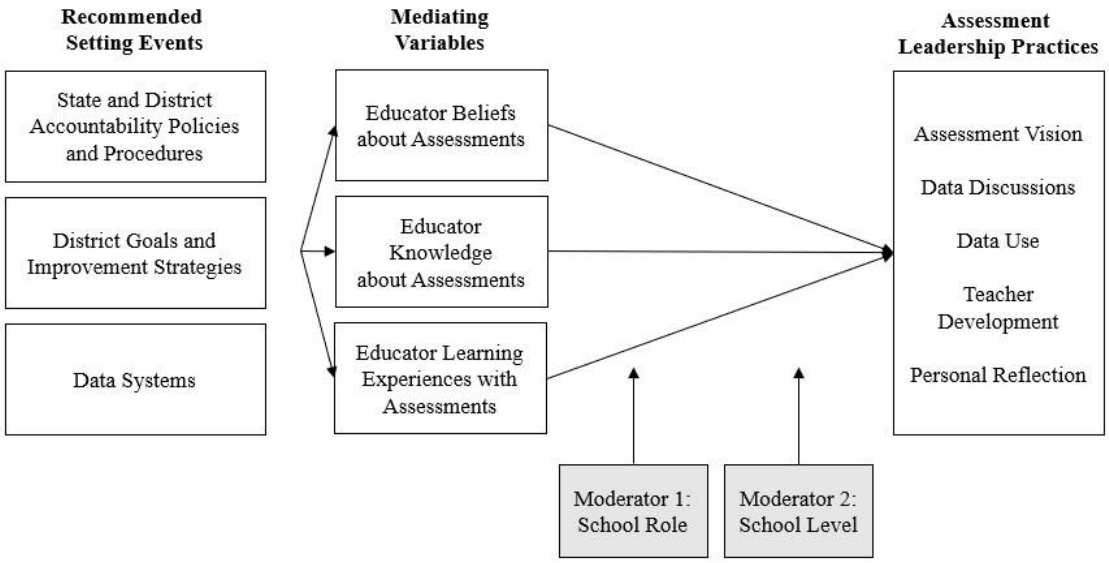

Figure 2.2. Key variables for study in Assessment Leadership.

Research questions. This study examined assessment leadership as a construct by measuring key assessment leadership variables that influence assessment practices in schools and classrooms. Figure 2.2 illustrates the components of the conceptual framework that were investigated in this study: recommended setting events, assessment learning experiences, beliefs, knowledge, and leadership practices. Three research questions were developed to explore the assessment leadership framework.

> RQ1. To what degree do assessment learning experiences, beliefs, and knowledge influence the assessment leadership practices of school and classroom leaders?

RQ2. To what extent does school role (i.e., school or classroom leader) moderate the relationship between assessment learning experiences, beliefs, and knowledge and assessment leadership practices of school and classroom leaders?

RQ3. To what extent does school type (i.e., elementary or secondary) moderate the relationship between assessment learning experiences, beliefs, and knowledge and assessment leadership practices of school and classroom leaders?

Summary

Chapter 2 provided theoretical foundations for the conceptual framework in assessment leadership, which incorporates critical school and classroom leader behaviors pertaining to student assessment systems and data-informed decision-making strategies as core components of school improvement. The need for continued research in assessment literacy and leadership were highlighted, specifically to understand (a) the relationships among leader experiences, beliefs, and knowledge; (b) their relative influences on school and classroom leader practices; and (c) influence school level type and educator role (i.e., administrator or teacher) has on the interactions among these variables. The literature review outlined key concepts for the basis of research design and data collection and analysis procedures discussed in Chapter 3 and establishes the foundation for analyzing and interpreting the results presented in Chapter 4 and discussed in Chapter 5.

CHAPTER 3

METHODOLOGY

The purpose of this study was to explore assessment leadership as a construct among school and classroom leaders in one large district in Florida. Data were collected using an Internet-based survey constructed from existing qualitative and quantitative measures of key components of assessment leadership established in the literature. The survey was constructed in Qualtrics and administered to the target population via district electronic mail. Survey items consisted of multiple-choice prompts to gather background and demographic information, Likert-type items adapted from a principal reflection tool (Noonan & Renihan, 2008) and items from an established tool entitled *Assessment Literacy Inventory* (Mertler & Campbell, 2005). Documents obtained from the target district that described the assessment-related and PLC activities over the previous four years also were obtained and reviewed. The research design allowed for standardized measurement of key assessment leadership variables, while framing the results in the context of one school district. The study is considered exploratory since components of the survey instrument were constructed and not already validated in prior studies.

Target Population

This study was conducted in the selected large school district in Florida for two primary reasons. First, the size of the district provided a sufficient participant pool to sample. At the time the district agreed to serve as the study site, the most recent data indicated it served 71,690 students across elementary and secondary grades. Special populations included 15% students with disabilities, 4% students that spoke English as a second language, and 59% students who received free or reduced-price lunch.

Additionally, there were 10,167 school board employees of which 5,028 (49%) were instructional employees working in 89 public and charter school sites. School sites included 47 elementary schools, 15 middle schools, 13 high schools, 3 education centers, 10 charter schools, and 1 virtual school. Of the 316 administrators in the district, 230 (73%) were school-level administrators. Second, during the past decade, school and classroom leaders in the selected district had participated in a series of professional development activities pertaining to standards-based instructional approaches and student assessment systems implemented within a professional learning communities (PLCs) model. The assessment and PLC reform efforts implemented in the district were aligned with the problem of practice and research questions for this study.

Given the scope of the study and size of the district, a comprehensive sampling method was used, wherein all school and classroom leaders were selected to participate (McMillian & Schumacher, 2010). School leaders were defined as school-level administrators (SLAs), both principals and assistant principals, who worked at elementary, middle, or high schools or who served at specialty schools (i.e., education centers, virtual school). Classroom leaders were defined as PLC facilitators selected by school leaders to receive additional professional development and facilitate PLC discussions among teacher teams at the school level. They were differentiated from other teacher leaders because their human resources designation indicated they receive the school district's PLC facilitator supplement, a stipend for their unique school role. The final participant pool was extracted from the district file based on the target population definitions for this study; that is, employees who did not receive the PLC facilitator supplement were not included in this study. In addition, school and classroom leaders at

the 10 charter school sites were excluded as study participants because they did not participate in district initiatives or receive the PLC facilitator stipend.

Table 3.1

Number of Individuals in the Target Population

School Role	Elementary	Secondary	Total
SLA-Principal	50	32	82
SLA-Assistant Principal	53	95	148
PLC facilitator	490	496	986
Totals	593	623	1,216

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

In all, the target population included 1,216 school board employees with 230 SLAs (i.e., principals, assistant principals) and 968 PLC facilitators across the 79 non-charter school sites. Table 3.1 displays the breakdown of participation candidates by school type and role with totals for each group. SLAs represented 19% of the target population while PLC facilitators represented 81% of the target population. Forty-five percent (45%) of the SLAs were at the elementary level with 55% at the secondary level. Fifty percent (50%) of the PLC facilitators were at the elementary level and 50% at the secondary level. The PLC facilitators identified for this study represented 24% of the total instructional employee population in the target school district.

Instrumentation

An Internet-based survey was constructed by the researcher to gather information on assessment leadership knowledge, experiences, and practices as well as on respondent demographics, herein called the *Assessment Leadership Survey* (ASLS) (see Appendix B). The ASLS survey consisted of 80 questions, organized into five sections: (1) educational background and experiences, (2) assessment beliefs, (3) assessment practices, (4) assessment knowledge as measured by the *Assessment Literacy Inventory* (ALI), and

(5) demographic questions. Section 4 contained items extracted from an existing online survey; the remaining items were developed from constructs in the established literature as critical factors to assessment leadership. The survey was created in Qualtrics and the items were randomized within and across scales to reduce order effects. The order of the sections was preserved to maintain consistency in presentation across respondents. In addition, the survey was designed to be completed on a computer or mobile device.

Prior to dissemination, input on item construction and administration was obtained from experts in measurement as well as educators in the field. The survey was reviewed by two university faculty and completed by several educational leaders skilled in survey development and classical item analysis. A few adjustments were made to the items to capture the elements of each of the variables (i.e., beliefs and practices) more effectively as well as to improve readability and ease of administration. The survey was then field tested with nine aspiring leaders in education programs at the University of Kentucky. The purpose of the field test was to (a) gauge the length of time to complete the survey and (b) solicit feedback on clarity of the items. Among the field testers, the survey was completed within the estimated 30 to 60 minutes; thus, no items were removed or altered from the survey. Minor item format issues were reported by field testers and resolved to improve administration on mobile and computer devices.

Educational Background and Experiences Section

Section 1 of the survey contained questions about school and classroom leader background and educational experiences. Section 1 was coded as BAC in the final sample analysis. Specifically, respondents were asked to identify their current school type (i.e., grade levels taught or lead) and role (i.e., principal, assistant principal, PLC

facilitator) in the school district and to report on gender, age (in years), years in education as a teacher, years in education as an administrator, total number of assessment courses in undergraduate and graduate programs, and estimated number of hours in professional development workshops pertaining to assessments that have been provided by the school district or attended in other continuing education situations such as at conferences. This responses from this section were used to describe the population as well as to explore the assessment learning experiences portion of the assessment leadership model.

Beliefs and Practices Sections

Sections 2 and 3 of the survey contained items that captured the assessment beliefs and practices components of the assessment leadership model (Noonan & Renihan, 2006). Section 2 was coded as BEL and Section 3 as PRA in the final sample analysis. A review of literature in assessment leadership evidenced vast gaps in measurement of these components: No known or accessible studies of assessment leadership with empirically validated outcomes were located. Most studies have relied on primarily qualitative approaches or reviews of the literature with little investigation into the variables that contribute to assessment leadership competencies and practices (Noonan & Renihan, 2006). Studies that have employed quantitative measures either used surveys without sample reliability statistics or incorporate factor analyses to adequately evaluate the survey (Carr, 2002; Hameister, 2013; Henry, 2011) or employed measures that are not readily accessible except through state departments or for purchase from companies (Matthews, 2007; Sterrett, 2005).

Assessment literacy measures have been established in the literature; however, they only capture one component of the assessment leadership model. The most

comprehensive known survey in assessment leadership is a 17-item principal self-reflection tool developed by Noonan and Renihan (2006) and organized into three key categories: knowledge, appreciations, and skills. *Knowledge* reflected competencies required to identify, use, and interpret appropriate assessment tools to inform decisions. *Appreciations* encompassed beliefs about assessments and data use for instructional decisions. *Skills* comprised the specific practices school and classroom leaders engage in as leaders and facilitators of assessment activities at school and classroom levels. While the tool was developed to promote principals' self-reflection of their own assessment leadership practices, items also were indicative of recommended models of assessment leadership established in other studies (Noonan & Renihan, 2006, 2010).

For this reason, coupled with the paucity of research on assessment leadership instruments, the principal self-reflection tool was used to construct the assessment beliefs and practices potion of the survey (Noonan & Renihan, 2006). The assessment beliefs section contained 14 items measuring the perceived degree to which participants agreed with the statements on a 4-point Likert-scale ranging from 1= *Strongly Disagree* to 4=*Strongly Agree.* The assessment practices section contained 18 items measuring the reported frequency of engagement in assessment leadership practices on a 5-point Likert-scale ranging from 1=*Never* to 5=*Almost Always.* Since responses to items in the second and third sections of the survey were self-reported by respondents and had not been empirically-validated, these 32 items served as a pilot for future research in this area.

Assessment Literacy Inventory Section

Section 4 of the survey contained a measure of assessment literacy entitled the *Assessment Literacy Inventory* (ALI) by Mertler and Campbell (2005). Section 4 was coded as ALI in the final sample analysis. ALI was designed to measure the assessment knowledge component of assessment leadership (see Figure 2.2), which is a measure of key assessment competencies required of assessment leaders. The 35 items were presented in a series of five classroom assessment scenarios whereupon survey respondents read a scenario and then responded to seven questions aligned with designed assessment standards. The ALI was scored for correct (1) or incorrect (0) responses. Totals and percentages were calculated for the seven standards as well as for the composite score for the complete inventory. Means and standard deviations were obtained for items as well as total scores as measures of central tendency.

ALI is the third iteration of an assessment literacy measure derived from assessment literacy principles established by the *Standards for Teacher Competence in the Educational Assessment of Students* to address growing concerns around teacher competencies in assessment literacy (AFT, NCME, & NEA, 1990). According to the standards, teachers should be skilled in (a) choosing assessment methods appropriate for instructional decisions; (b) developing assessment methods appropriate for instructional decisions; (c) administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods; (d) using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement; (e) developing valid pupil grading procedures that use pupil assessments; (f) communicating assessment results to students, parents, other lay

audiences, and other educators; and (e) recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information. These seven competencies were measured across five classroom-based scenarios.

A review of empirical studies suggested that the different iterations of the ALI, based on these standards, have been the most widely used tools for capturing assessment literacy levels of principals and teachers at both preservice and inservice levels, based on test content from the *National Council on Measurement in Education* (Mertler, 2003). The ALI is an adaptation of the *Classroom Assessment Literacy Inventory* (CALI) by Mertler (2003), which was derived from the *Teacher Assessment Literacy Questionnaire* (TALQ) developed by Plake (1993). The TALQ was the first to measure teachers' assessment literacy levels using all seven principles. The studies showed differences in overall reliability coefficients of inservice teachers at KR20 = .54 (Plake, Impara, & Fager, 1993) compared to preservice teachers at KR20 = .74 (Campbell, Murphy, & Holt, 2002), KR20 = .84 (Alkharusi et al., 2011), and KR20 = .77 (Beziat & Coleman, 2015).

CALI was refined to incorporate more clear scenarios with the intention of improving the technical adequacies of the survey (Mertler & Campbell, 2005). Several items were revised following pilot phases. Item analyses resulted in the removal of some items from the CALI, which yielded a higher estimated reliability coefficient for preservice teachers (KR20 = .75). Across studies and survey versions, it was evident that ALI was useful in measuring the assessment literacy levels of principals and teachers. However, it also was evident that more research is necessary to validate further the results of the survey, such as using it to evaluate the assessment literacy skills of school and classroom leaders engaged in professional development activities in this area.

In prior studies, results have suggested differences among the seven scales with strengths in knowledge of administering and scoring assessments and relative weaknesses in communicating the results (Plake et al., 1993). Other studies suggested distinctions among standards for using the assessment data for instructional decisions compared to recognizing unethical practices or when comparing principal and teacher responses (Perry, 2013) and differences between subject matter areas (Davidheiser, 2013). Although there have been evident differences among scales in some studies, others suggest that ALI can be treated as a unidimensional model based on factor loadings (Alkharusi et al., 2011). For these reasons, total score in addition to scale scores were computed to analyze differences among within assessment literacy factors.

Demographic Section

Section 5 of the survey contained questions pertaining to highest level of education, gender, ethnicity, and age, which were coded as DEM in the final sample analysis. The purpose of these questions was to gather descriptive information with respect to the demographics of the target population. The final question presented the proposed assessment leadership definition and inquired the extent to which the participants believed the definition matches their current role as school or classroom leader and whether the participants would rewrite the definition. The intention of this question was to collect additional context for analysis and to inform future research.

Data Collection

First, a review of the districts policies, procedures, and professional development activities pertaining to assessment-related and PLC initiatives was conducted. Documents were selected and sent from leaders in the departments of Professional

Development and Assessment, Research, and Measurement within the district. These documents encompassed descriptions and exemplars of implementation of assessment-related and PLC initiatives involving school and classroom leaders over the past couple of years. The documents provided context concerning how this district has approached implementation of their assessment-related and PLC initiatives.

Second, district electronic mail addresses for the target population were obtained from the Human Resources Department in the Florida school district selected as study site. An invitation to complete the survey was sent to potential participants in a personalized electronic-mail link executed in Qualtrics. The message contained the first paragraph of the consent form as a prompt, followed by a link to the Internet-based survey. Participants were notified they could complete the survey on their mobile device or computer; while it was recommended to complete the entire survey at one time, they could return to the last answered question to complete the survey at a later time.

The electronic-mail link opened to an electronic cover letter that described the study and articulated directions for completing the survey, including the estimated completion time of 30 to 60 minutes. Given the comprehensive nature of the survey and the depth of knowledge required to respond to the assessment literacy items, participants were encouraged to answer each item to the best of their abilities at the time. They were informed that all responses would be secured to protect confidentiality, and they had the option to discontinue participation at any time. Before proceeding with the survey, participants were asked to verify informed consent at the bottom of the cover letter.

The target sample size was set at 600 or approximately 50% of the target population. The survey was initiated near the end of March 2017 and left open for six

weeks. The time period was supported by district leaders because it was as occurring after Year 4 implementation of their initiatives but prior to administration of state assessments of student learning. To meet the target sample size, weekly reminders were emailed to study participants who had started but had not completed the survey and to qualified participants who had not yet started the survey.

Missing Data

Missing data are an unfortunate reality of survey research. In fact, rates of 15% to 20% missing data often are observed in educational research (Enders, 2003) and have become a common problem in large-scale survey research (Peng, Harwell, Liou, & Ehman, 2006). Even though missing data are expected in survey research, it can compromise statistical power and generalizability as well as introduce more standard error than likely observed in a complete dataset, especially when missingness exceeds 5% of the total responses (Brick & Kalton 1996; Cheema, 2012; Dong & Peng, 2013; Enders, 2010). Fortunately, in the recent decades, approaches to handling missing data have been well researched with guidelines based on the observed patterns of missing data.

Most studies handle missing data use listwise deletion (LD) or pairwise deletion (PD) methods (Brown, 2015; Enders, 2010; Manly & Wells, 2015, Peugh & Enders, 2004). LD is the simplest approach as it removes all cases with missing data and has shown to be a viable method for handing missing data (Cheema, 2012). However, LD alone greatly reduces sample size and increases nonresponse bias, errors resulting from the differences between the responses of the study participants who completed the survey and those who did not (Brick & Kalton, 1996; Garson, 2015; Enders, 2010). Moreover, while most frequently used, both LD and PD have been shown to be less robust in

eliminating bias than newer, principled methods for handling missing data (Acock, 2005; Brick & Kalton 1996; Dong & Peng, 2013; Enders, 2010).

Principled methods include multi-imputation (MI), full information maximum likelihood (FIML), and expectation-maximization (EM) method are preferred to Listwise Deletion (LD) and Pairwise Deletion (PD) methods (Brown, 2015; Dong & Peng, 2013; Enders, 2010). These methods "do not replace the missing value directly," but instead, "combine available information from the observed data . . . in order to estimate the population parameters and/or the missing data mechanism" (Dong & Peng, 2013, p.1). These methods assume multivariate normality with MI less likely to violate normality than FIML and EM (Schafer, 1997). Intuitively, the greater percentage of missing data in a dataset, the more threats to reliability and validity (Dong & Peng, 2013). However, researchers agree the way missing data are handled is more critical than the percentage of missingness (Enders, 2010; Wainer, 2010). While there are limitations, statistically supported analyses can be accomplished, even with significant portions of missing data (Pampaka, Hutcheson, & Williams, 2015).

Fortunately, several methods for handling, analyzing, and reporting on missing data have been established in the literature (Acock, 2005; Brick & Kalton, 1996; Brown, 2015; Enders, 2010; Dong & Peng, 2013; Johnson & Young, 2011; Liu & De, 2015; Manly & Wells, 2015; Pampaka, Hutcheson, & Williams, 2015; Schafer & Graham, 2002; van Buuran, 2007). Across methods, researchers generally recommend following three key steps: (1) examine missingness, (2) determine how missing data will be handled, and (3) proceed with analysis using the most appropriate missing data methods. Given the nature of missingness in this study, best practices for handling missing data in

survey research were applied in two phases: first to the original sample and then to the final sample. Response rates were reported for the original and final samples.

Phase 1: Original Sample

Sources of missing data typically fall into four categories: non-coverage, total nonresponse, item nonresponse, and partial item nonresponse (Brick & Kalton, 1996). Noncoverage occurs when a faction of the target population is not represented in the sampling population and thus do not have the opportunity to respond. Total nonresponse occurs when a select group of respondents from the sampling population do not respond to any of the items. Item nonresponse occurs when respondents do not answer a few items. Partial item nonresponse occurs when respondents stop responding and a portion of the survey is incomplete. Upon review of the original sample in this study, the sources of missing data spanned total, item, and partial item nonresponse categories, depending on the degree of missingness for each respondent. Across respondents, approximately 19% contained total nonresponses, 11% contained item nonresponses, and 50% contained partial item nonresponses.

The original sample was revised by deleting cases with total nonresponses and large portions of partial item nonresponses. First, total nonresponses (i.e., respondents who clicked they agreed to the consent form but did not answer any of the items) were removed from the sample. Second, respondents who provided their school role and type, but did not provide responses for any of the assessment leadership variables in the study (i.e., postsecondary courses, professional development sessions, assessment beliefs, assessment practices, assessment knowledge) were removed from the sample. Last, additional respondents who did not indicate their role or report years of service for school

or classroom leader (i.e., only reported for years for classroom teacher) were removed. Two-hundred eighty-four (284) total respondents were retained in the final sample. Even though this approach reduced the sample size by 28%, the final sample contained complete categories for school type (i.e., grade levels) and role (i.e., principal, assistant principal, PLC facilitator); years of experience as a classroom teacher, PLC facilitator, and administrator; and frequency of PLCs (i.e., weekly, bi-weekly, monthly, quarterly). Dong and Peng (2013) assert that retaining as many complete categories as possible improves statistical power when handling datasets with large amounts of missing data.

Phase 2: Final Sample

Enders (2010) outlined six patterns of missing data in survey research: univariate, unit nonresponse, monotone, general, planned, and latent. Univariate patterns contain missing data isolated to a single variable. Unit nonresponse patterns have missing data for portions of a survey (i.e., more than one variable). Monotone patterns are observed when respondents drop out and do not return to complete. General patterns occur randomly throughout the dataset and do not appear to be correlated with a specific variable. Planned patterns happen when researchers intentionally distribute only portions of the survey to decrease number of items respondents must complete. Latent patterns are when latent variable values are missing for the entire sample. The distribution of observed to missing data for the surveyed items (not the computed items) were examined using IBM®SPSS® missing data descriptives. Figure 3.1 reflects the percentage of complete and missing data by variables, cases, and values. As shown, 92% (n = 76) of the variables contained at least one missing value and 79% (n = 224) of the cases had at least one missing value. The variables that did not contain missing data were constructed

in Phase 1 to create complete data categories in preparation for imputation (Dong & Peng, 2013). In the final sample, 36% of the total values in the dataset were incomplete, which was improved from 51% in the original dataset.

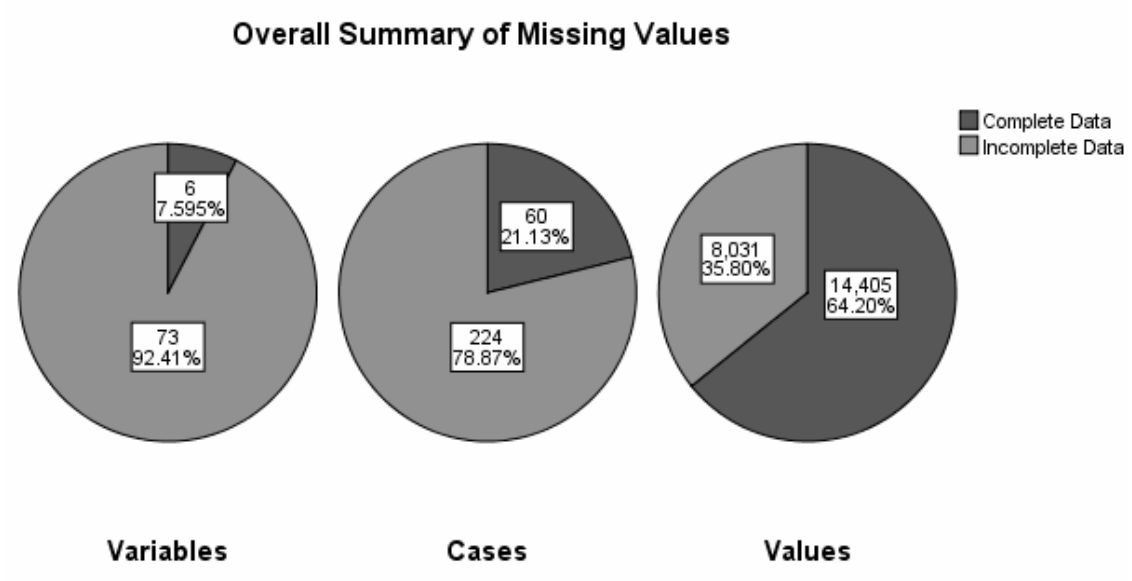

Figure 3.1. Percentage of missing values in the final sample.

Figure 3.2 illustrates the missing value patterns in the final sample. It was evident that some respondents who completed the survey skipped items while other respondents both skipped items and failed to complete survey items after certain points, mostly notably at Section 4 (ALI). These patterns suggested both general and monotone tendencies, as explained in more detail below. Additional analyses were conducted to further understand how missing data were dispersed among sections and variables.

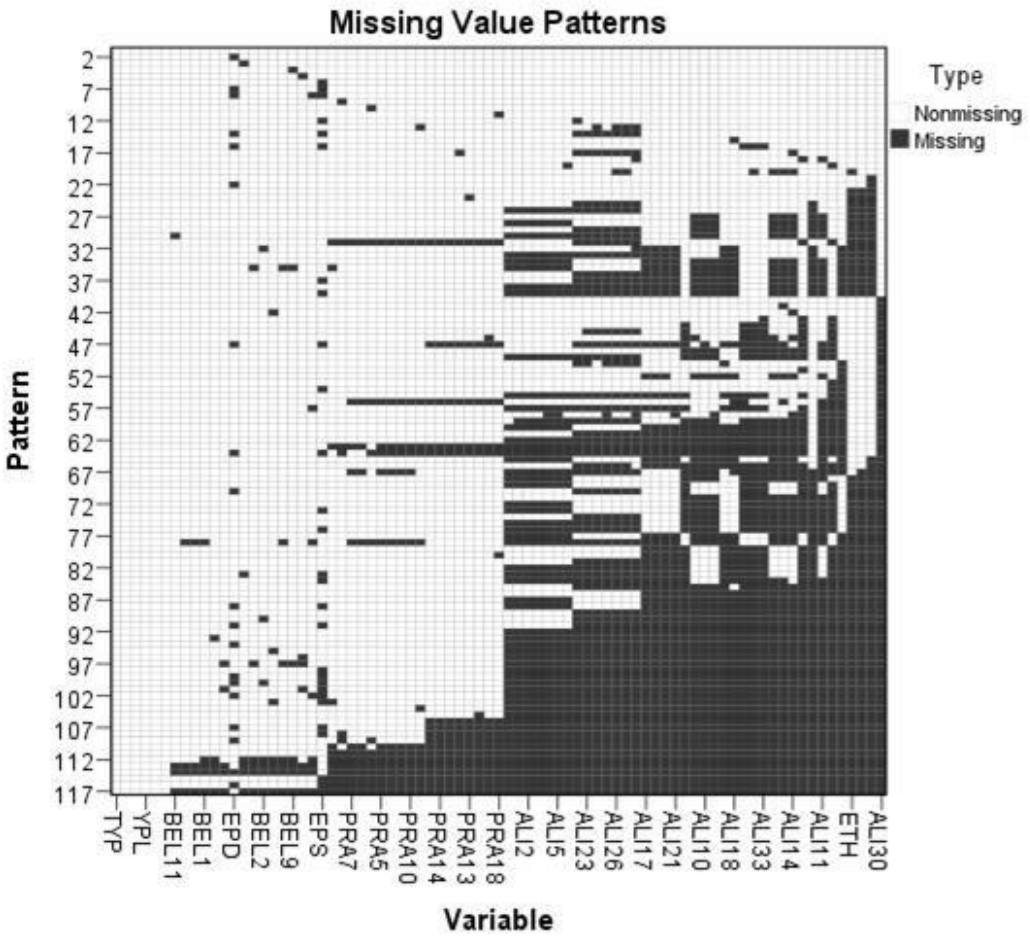

Figure 3.2. Patterns of missing values in the final sample.

As shown in Table 3.2, the percentage of missing data in the survey ranged from 9.9% in Section 1 to 61.6% in Section 4. The counts and frequencies by section and successive items suggested an increasing amount of missing data as respondents progressed through the survey. Most respondents completed items through Section 3 when total survey completeness was reduced by almost half. This tendency resulted in significantly increased missing data for the ALI (i.e., assessment knowledge) compared to the other study variables. Even though the attrition after Section 3 of the survey suggested a monotone pattern, it was evident that missing data were not purely monotone, but also contained some general missing data patterns (Enders, 2010). Given

the presence of monotone patterns, without removing additional survey responses, it was not possible to reorder the cases into a purely monotone pattern; thus, the dataset was treated in analysis as having general patterns of missingness (van Buuran, 2007).

Table 3.2

Patterns of Missing Data in the Final Sample

	Ranges of complete data		Ranges of missing data	
Survey Item(s)	*f*	%	*f*	%
Section 1: Background	245 – 284	86.3 – 100	0 – 39	0 – 13.7
Section 2: Beliefs	254 – 258	89.4 – 90.8	26 – 30	9.2 – 10.6
Section 3: Practices	210 – 218	73.9 – 76.8	66 – 74	23.2 – 26.1
Section 4: Knowledge	109 – 129	38.4 – 35.4	155 – 175	54.6 – 61.6
Section 5: Demographics	111 – 118	39.1 – 41.5	166 – 173	58.5 – 60.9
Postsecondary courses	245	86.3	39	13.7
Professional development	256	99.6	28	9.9
Beliefs total score	238	83.8	46	16.2
Practices total score	197	69.4	87	30.6
Knowledge total score	78	27.5	206	72.5

Missing data patterns described how values were dispersed across the dataset; however, they did not explain why values are missing. Rubin (1987) identified three missing data mechanisms: missing at random (MAR), missing completely at random (MCAR), and missing not at random (MNAR). MCAR assumes missing data are independent of the observed and missing responses with an equal likelihood of missingness in the dataset. The missingness are not related to the data itself (Enders, 2010). MAR assumes missing data are independent of the missing responses but may be dependent on other observed responses. The probability of missingness is related to at least one other variable in the dataset (Enders, 2010). Last, MNAR assumes missing data are dependent on both observed and missing responses, which suggests

omissions are contingent on variables not inherent in the dataset. Identifying the missing data mechanism is critical to selecting the most appropriate method for handling the data.

Unfortunately, satisfying the missing data mechanism given the observed patterns of missing data can be challenging (Enders, 2010; Schafer & Graham, 2002). While there are limitations, some approaches have been established in the literature (Garson, 2015). For this study, Missing Value Analysis (MVA) in IBM®SPSS® was conducted using observed variables. The expectation maximization (EM) algorithm with Little's chi-square statistic was computed assuming normal distribution and maximum iterations of 5000. Separate variance t-tests calculations were run for scale variables while crosstabulations were run between categorical and scale variables. These data were used to further examine the underlying assumptions of missingness.

Little's chi-square statistic was significant at $\chi^2(df = 5{,}846, n = 284) = 6128.4$, $p = .01$, which rejects the null hypothesis that missing data are MCAR. Separate variance *t* tests suggested several variables yielded significant values, which further supported the theory that data are not MCAR. For example, administrators with fewer years of experience tended to have more missing values for postsecondary courses, $t(244) = 7.42$, $p < .001$, and professional development sessions $t(39) = 2.13$, $p = .04$, then administrators with more years of experience. Moreover, cross-tabulations of categorical and indicator variables for school type and school role suggested minimal differences less than 5%. Given the outcomes of these three tests, some relationship between the observed and missing variables likely was present. Thus, the mechanism for missing data was assumed to be MAR and not MCAR or MNAR.

Multiple imputation (MI) has been suggested as the standard for handling MAR mechanisms that follow a general missing data pattern (Allison, 2001; Dong & Peng, 2013; Enders, 2010; Garson, 2015; Schafer & Graham, 2002; van Buuran, 2007). MI uses all available data to create multiple complete datasets by making statistical inferences about the missing data as opposed to merely calculating a mean of the observed values. Even though MI has its limitations, this method often is preferred over other principled methods in that it has various software options, is applicable to a broader range of statistical models, is superior for categorical variables, and estimates as if results are derived from complete datasets, which allows for expanded analysis techniques (Johnson & Young, 2011; Liu & De, 2015; Pampaka, Hutcheson, & Williams, 2015; van Buuran, 2007). MI also is more widely accepted for handling any missing data pattern since assumptions often are hard to verify and contain inherent bias (Enders, 2010; Garson, 2015). MI tends to be less likely to experience convergence issues than FIML and EM when multi-variate normality is violated. Given all these factors, and the conditions of the dataset in this study, MI was determined as the best method.

Two MI models typically are used: multivariate normal imputation (or joint modeling (JM) and fully conditional specification (FCS) (Allison, 2001; Acock, 2005; Dong & Peng, 2013; Enders, 2010; Johnson & Young, 2011; Liu & De, 2015; Manly & Wells, 2015; van Buuran, 2007). The JM imputation model assumes joint multivariate normal distribution of all variables in the dataset (Allison, 2001; Enders, 2010). For JM, imputations are created based on the pre-specified distributions and are suitable for continuous variables only. The FCS model assumes more complex relationships among variables that may not be accounted for in JM (van Buuran, 2007). For FCS, imputations

are created based on iterative sets of regression equations that are appropriate for each variable and regressed on all the other variables in the dataset. Due to its flexibility in predicting missing data, especially if the wrong assumptions have been identified, FCS MI has been purported as the best method and model for handling missing values for categorical and continuous variables and, thus, was used in this study (Enders, 2010; Johnson & Young, 2011, Liu & De, 2015; Manly & Wells, 2015; van Buuran, 2007).

MI FCS contains a three-step process: imputation, analysis, and pooling (Enders, 2010). Generally, it is recommended that the minimum number of imputations account for the percentage of missingness in the original dataset (Manly & Wells, 2015; White Royston, & Wood, 2011). However, to enhance statistical power, particularly for studies using varying statistical methods, imputations larger than percent missing should be calculated (Dong & Peng, 2013; Enders, 2010). While there are differing views as to whether to include dependent and interaction variables in the imputations, consensus is to include all variables in the dataset to reduce biases inherent in those variables at the analysis stage (Allison, 2001; Garson, 2015; Graham, 2009). Excluding analysis variables even under MCAR or MAR conditions could weaken associations among other variables (Enders, 2010). Thus, MI FCS was conducted for 100 imputations using Markov Chain Monte Carlo (MCMC) convergence with 10 iterations and the maximum number of model parameters set at 5,000.

Diagnostics between the observed and imputed values were conducted on the imputed dataset to ensure the multiple imputations were reasonable (Liu & De, 2015; Manly & Wells, 2015; White Royston, & Wood, 2011). Since the worst linear function (wlf) statistic was not available in IBM®SPSS®, convergence was examined by plotting

the means and standard deviations by iteration and imputation for each scale dependent variable on a line graph. Graphical patterns should appear random with few discernible trends (Enders, 2010). Graphs were constructed and analyzed for independent variables of beliefs (BEL), experiences (EPS and EPD), and knowledge (ALI) as well as for dependent variable of practices (PRA). Figures 3.3 through 3.8 illustrate the MI FCA convergence patterns by variable for each of the 100 imputations across the 10 iterations.

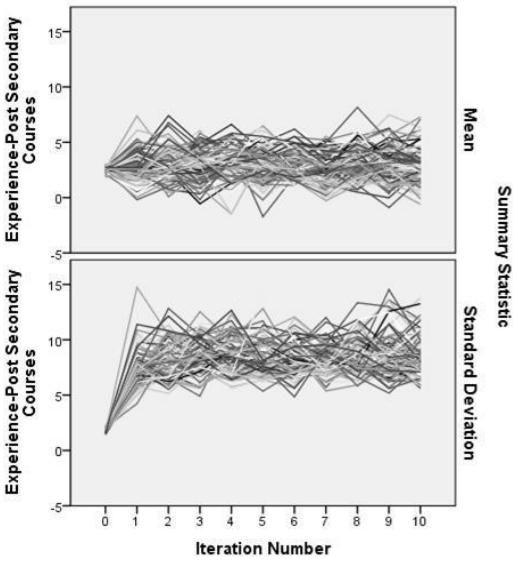

Figure 3.3. EPS at 10 iterations.

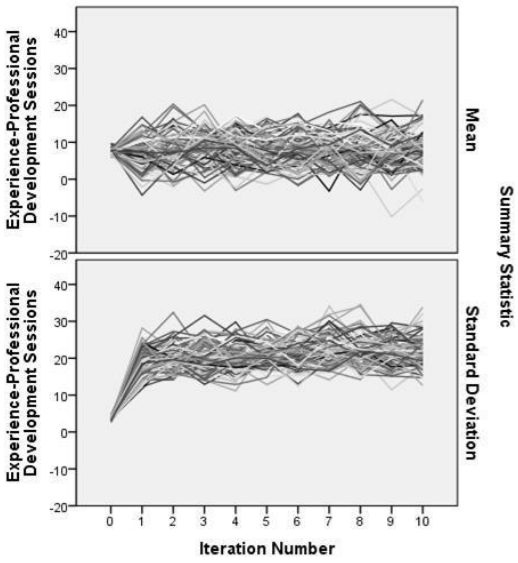

Figure 3.4. EPD at 10 iterations.

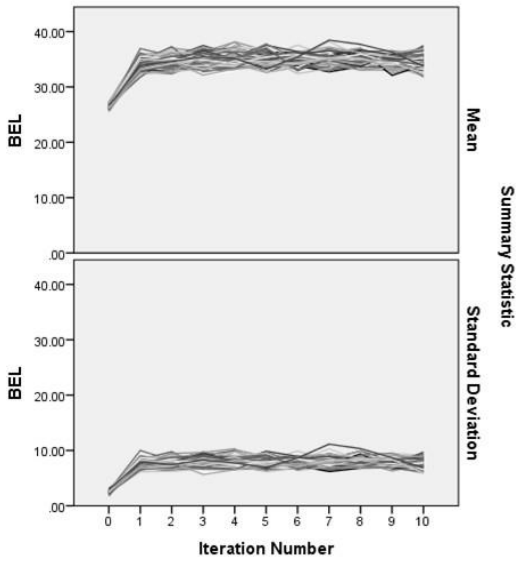

Figure 3.5. BEL at 10 iterations.

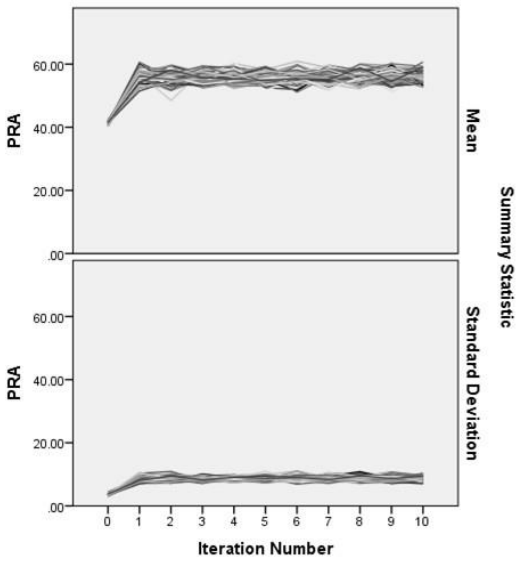

Figure 3.6. PRA at 10 iterations.

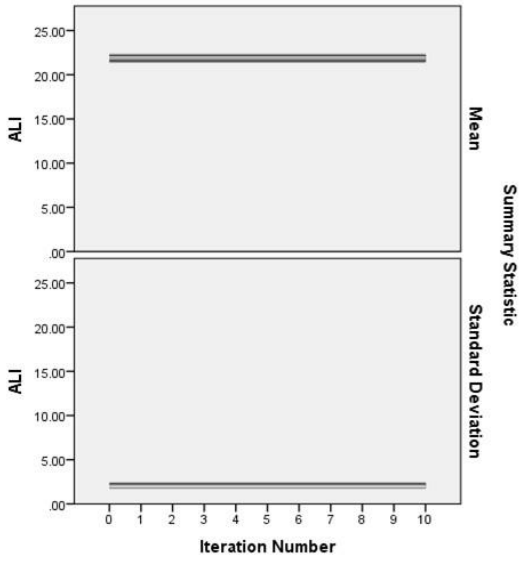

Figure 3.7. ALI at 10 iterations.

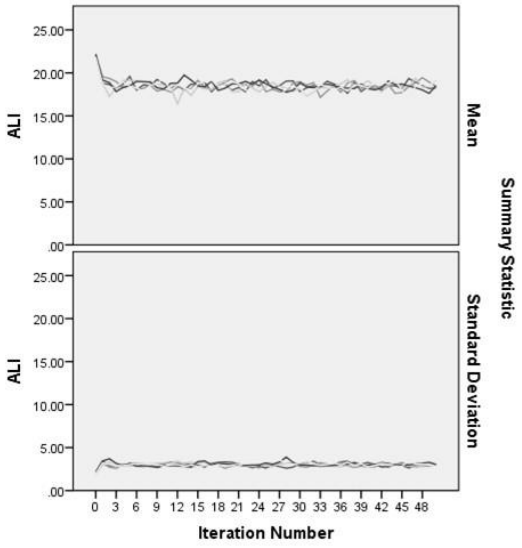

Figure 3.8. ALI at 50 iterations.

As shown, FCS convergence plots for EPS, EPD, BEL and PRA variables did not show a discernable trend. The means and standard deviations alternated randomly, which suggested they converged to a stable distribution (Enders, 2010). Different than the other variables, convergence problems were evident for ALI. This outcome is not surprising, given the significant increase in missing data for ALI compared to the other variables.

The best approach to resolve convergence issues is to increase the number of maximum iterations (Enders, 2010). Thus, MI FCS was rerun for ALI with 5 imputations at 50 and 100 maximum iterations. Figure 3.7 indicates improved convergence with the increase to 50 iterations and no noticeable difference at 100 iterations in Figure 3.8. Given these outcomes for ALI total score, MI FCS was executed again for all variables in the dataset at 100 imputations with 50 maximum iterations and 5,000 maximum model parameters.

Descriptive Analyses

Descriptive statistics were used to report respondent background information and demographics (McMillan & Schumacher, 2010). Total numbers and percentages were computed and reported by school role and type for years of experience in school and classroom leaders' roles, number of preservice undergraduate and graduate assessment courses, number of inservice professional development workshops and continuing education experiences, and current frequency of PLC meetings in the target school district. Demographic data also were calculated based on the original dataset and compared to the total population sample. These data were used to describe the characteristics of the final sample, excluding the nonresponses and partial responses.

Means and standard deviations were computed from the 100 imputed datasets based on the pooled item means for assessment learning experiences (EXP), beliefs (BEL), practices (PRA), and knowledge (ALI). These data also were calculated for the total measure scores for the independent and dependent variables as well as for the seven ALI (assessment knowledge) standards. Results were reported by school role and school type. Two-tailed independent sample *t* tests were run using Levene's Test for Equality of Variances assuming equal variances. The purpose of these diagnostics was twofold: (1) to

describe the average pooled responses and (2) to examine any mean differences between groups for the outcome variables by school role and school type.

Factor and Item Analyses

Before testing the research hypotheses, a series of factor and item analyses were conducted to evaluate the dimensionality of the survey measures for assessment beliefs (BEL), practices (PRA), and knowledge (ALI). The purpose of this step was to identify the simplest and most robust factors that explain the relationships between the observed variables and the latent variables (Brown, 2015). Since IBM®SPSS® does not pool multiple imputations for factor analysis, each imputed dataset was saved as a separate data file in preparation for pooled analyses in *Mplus.* Due to significant missing data, multiple imputed datasets, and marginal to poor outcomes from goodness of fit index examinations, several approaches were taken to analyze the factor structures.

Confirmatory Factor Analysis

First**,** *Mplus* was used to conduct a confirmatory factor analysis (CFA) for the beliefs (BEL), practices (PRA), and knowledge (ALI) measures with the 100 imputed datasets (Muthén & Muthén, 2017). As noted, CFA with multiple imputations is preferred because other methods tend to underestimate variances and overestimate correlations among variables (Brown, 2015; Enders, 2010). The maximum likelihood parameter (MLR) was selected as the most appropriate estimator because it corrects for non-normality and best used with the MI FCS method missing data procedure employed in this study (Brown, 2015; Jackson, Gillaspy, & Purc-Stephenson, 2009). Since the BEL and PRA measures have not yet been established in the literature and were designed with the intension of measuring one construct, CFA was conducted assuming

unidimensionality for these measures. Alternatively, given the known factor structures of ALI in the literature, CFA was employed for ALI based on one and seven factor solutions, to attempt to fit for the seven ALI standards (Mertler & Campbell, 2005).

No single fit index has been established for CFA. Thus, several fit indices were evaluated: comparative fit index (CFI), Tucker-Lewis index (TLI), root-mean-square error of approximation (RMSEA), and standardized root-mean-square residual (SRMR). The fit outcomes were then compared to fit index cutoffs recommended in the literature (Jackson, Gillaspy, & Purc-Stephenson, 2009; Preacher & MacCallum, 2002). The unidimensional models suggested poor fit across indices for all measures, even when removing items with loading estimates below 0.4. Additional CFAs were employed to explore other possible factor structures. CFAs were run for two-, three-, and four-factor solutions as well as for the bi-factor solution. The bifactor model is an alternative approach to measuring multi-faceted constructs that accounts for both the general and the specific factors that may be present (Chen, Hayes, Carver, Laurencea, & Zhang, 2012). Even though the testlets were not intended to measure separate constructs, it was hypothesized the organization of items may have influenced the factor structure. Thus, the bifactor model would correct for possible error. Across multiple CFA outputs, fit indices suggested marginal fit at best, even with dropping items and forcing factors. Marginal to poor fit, even for ALI, likely can be attributed to small sample size and significant missing data (Brown, 2015). As a result, it was determined to conduct Exploratory Factor Analysis (EFA) to further examine the underlying relationships among measure variables. CFA and EFA together were used to inform and maximize decisions about factor structures (Gorsuch, 1988).

Exploratory Factor Analysis

Next, IBM®SSPS® was used to conduct exploratory factor analysis (EFA) for the beliefs (BEL), practices (PRA), and knowledge (ALI) measures. Principal Axis Factoring (PAF) was used to extract factors based on an inspection of the scree plot for the original dataset as well as for each of the 100 imputed datasets. Since IBM®SSPS® does not pool results for factor analysis, each individual imputation was reviewed and compared to the original dataset to determine the best decision for analysis. Several assumptions were examined for beliefs (BEL), practices (PRA), and knowledge (ALI) measures to ensure the final sample was suitable for factor analysis. Correlation matrices were reviewed for reasonable correlation coefficients between 0.3 and 0.9 with few significance levels greater than $p < .05$. Kaiser-Meyer-Olkin (KMO) and Bartlett's Test of Sphericity were calculated as measures of sampling adequacy. The original and imputed datasets were evaluated for KMO > 0.8 and Bartlett's as statistically significant at $p < .05$. Once assumptions were verified within range for BEL, PRA, and ALI, the measures were examined for number of factors.

Scree plots for the 100 imputed datasets were analyzed heuristically and compared to the original datasets to identify the most common number of factors for each measure. Across multiple imputations, the number of factors ranged from 2 to 4 for beliefs (BEL), 3 to 5 for practices (PRA), and 11 to 14 for knowledge (ALI). EFAs were rerun for beliefs (BEL) and practices (PRA) forcing 3 and 4 factor solutions with a direct oblimin rotation, respectively, which were the most common number of factors extracted across imputations. The results of the forced factor analyses did not yield interpretable patterns. There were no discernable patterns that supported discreet identifiable factors

within any of the three measures. Given this finding, coupled with the 3:1 ratio of the first eigenvalue to the second eigenvalue, it was determined to treat the measures as essentially unidimensional for analysis (Gorsuch, 1988).

EFAs were run again for beliefs (BEL) and practices (PRA) measures with forced factor of one. Items with factor pattern loadings less than 0.4 across the 100 imputed datasets were identified and then removed one at a time. EFA was rerun until a final factor structure was established. Unfortunately, the outcomes were less favorable for knowledge (ALI), likely due to limitations associated with significant degree of missingness and imputed values compared to the other measures. EFA was run for knowledge (ALI) assuming one, five, and seven factor solutions, given the recommended structures from the previous studies. The results of the EFA suggested unstable factor structure with low correlations. Finally, in order to retain some of the items for modeling purposes, a classical item analysis was conducted.

Classical Item Analysis

Last, classical item analysis was conducted in IBM®SSPS® for the 35-item ALI using the original and the 100 imputed datasets. This procedure assisted in identifying poor performing items on a scale by comparing individual item responses to the total test score (Brown, 2015). Reliability statistics were averaged across the imputed datasets and examined for coefficient alpha above .8. Corrected item-totals and coefficient alpha calculations also were averaged across ALI items. Items were selected for deletion if they improved the reliability above the average for the datasets. This technique was applied twice until the strongest internal consistency was obtained.

Inferential Analyses

The first research question (RQ1) was: *To what degree do assessment learning experiences, beliefs, and knowledge influence the assessment leadership practices of school and classroom leaders?* A multiple regression (MR) was conducted in *Mplus* using the reduced measures with the 100 imputed datasets to model the predictive nature of the independent variables (i.e. assessment learning experiences, beliefs, knowledge) on the dependent variable (i.e., assessment practices). Both maximum likelihood (ML) and maximum likelihood parameter (MLR) estimators were calculated. ML assumes multi-variate normal whereas MLR corrects for non-normality in the data (Enders, 2010). Since there was little difference in MR outputs, ML was used in the final analysis. The R^2 and *F* statistics were reported.

The second (RQ2) and third (RQ3) research questions were: *To what extent does school role (i.e., principal, assistant principal, or PLC facilitator) moderate the relationship between assessment learning experiences, beliefs, and knowledge and assessment leadership practices of school and classroom leaders? To what extent does school type (i.e., elementary or secondary) moderate the relationship between assessment learning experiences, beliefs, and knowledge and assessment leadership practices of school and classroom leaders?* Hierarchal multiple regression analysis was conducted to determine if the relative influence of assessment learning experiences, beliefs, and knowledge on assessment practices varied based on school role and school type. The full model was evaluated for significant interactions at $p < .05$. The R^2 and *F* statistics were reported. The difference in the R^2 estimates for the original and full models was

compared to determine whether the school role and school type moderators significantly changed the original model to further explain the relationship among the variables.

Summary

Chapter 3 presented the research design and the data collection and analysis methods. This study employed a non-experimental quantitative design to measure the relationships among assessment learning experiences, beliefs, and knowledge and the assessment leadership practices among school and classroom leaders. Survey data were collected using an Internet-based survey, with invitation to respond to survey sent through electronic mail to school and classroom leaders in one Florida school district. Due to significant missing data patterns, several steps were taken to analyze and prepare the data for analysis using known missing data procedures. The final sample data were analyzed using various descriptive and inferential statistical techniques and explained within the context of the assessment policies, practices, and professional development activities that are currently implemented in the target district. The results, explained in Chapters 4 and discussed in Chapter 5, are intended to extend the research base in assessment leadership constructs and measurement tools as well as contribute to the field of practice in assessment reform and models of instructional leadership.

CHAPTER 4

RESULTS

This study examined assessment leadership as a construct by investigating the relative influence of assessment learning experiences, beliefs, and knowledge on assessment practices in school and classroom leaders. A non-experimental correlational research design was employed to measure assessment leadership using an Internet-based survey constructed from a principal reflection tool (Noonan & Renihan, 2008) and an established assessment literacy inventory (Mertler & Campbell, 2005). The survey was administered to school and classroom leaders in one school district in Florida. The target school district was selected based on its implementation of assessment-related and PLC initiatives over the last four years. Descriptive and inferential analyses were conducted to examine relationships among variables in response to the research questions in this study.

Target Population Description

As noted, the target school district is a large district in Florida with over 70,000 students and over 10,000 school board employees across 89 school sites, including charter schools. The district has implemented several school reform initiatives in response to policy and practice shifts in the past four years. A comprehensive document review was conducted as part of this study to understand the content and scope of school reform efforts, specifically as it pertains to assessment-related and professional learning community activities in the past four years of implementation. The review included an overview of Florida state statutes pertaining to student assessment.

School District Context

Like other states, student assessments have been significantly influenced by state policies in the last decade. These policies have shaped local actions for school reform and respective school district initiatives. Specifically, Florida has not only required local districts to administer annual statewide assessments at specific grade levels, but also has required students in third grade, eighth grade, and approaching graduation to pass examinations with specific proficiency levels to proceed to the next grade level or graduate. These restrictions elevated the stakes compared to other states that use statewide assessment data to assign school grades, but not to prevent students from progressing to the next levels. In recent years, Florida state statutes have required local school districts to administer end-of-course assessments in every course, which are intended to measure progress toward state standards. However, they have raised the bar for school and classroom leaders to develop sound assessments that both match the rigor of the academic standard and reliably measure student-learning outcomes for every course, including electives.

In the target district for this study, the continuous improvement process centered on three main principles of practice: high impact instruction, data-driven decisions, and collaborative culture. Each principle was attached to a clear goal, executed in a step-wise fashion throughout the school year, beginning with success planning. The goals were aligned with the common vision of instructional excellence, which encompassed building strong content knowledge, thinking critically, collaborating and communicating, utilizing a variety of resources, and student taking ownership of their learning. To meet the vision, schools were provided with district resources based on need and informed by student data

at the start of each school year. Student-learning outcomes were continuously monitored using a variety of quantitative and qualitative data sources including district-led walkthroughs. Mid-year, a comprehensive need assessment (CNA) was conducted at each school using student data coupled with stakeholder feedback and self-assessment input. The CNA measured instructional leadership, high-impact instruction, collaborative culture, and data-driven decision making as core functions of school and classroom leaders. The CNA was utilized to evaluate implementation outcomes and return on investment for current efforts as well as to plan for continuous improvement in future implementation years. Resources included consultative support by experts, more time in PLCs, instructional demonstration videos and student work samples, and other tools.

Several professional development sessions were implemented to assist administrators and teachers to integrate policy into practice. Beginning in Fall 2013, team-lead supplements were repurposed to supplement teachers as professional learning community (PLC) facilitators at elementary and secondary schools. Administrators and PLC facilitators were provided with a series of quarterly professional learning experiences, facilitated by the school district and intended to establish standardized approaches to implementation across schools in the Florida district. District leaders studied the impact of training outcomes using walk-through guides and other tools tied to the objectives of the schoo- reform initiatives such as used in the CNA. Continuous adjustments were made in response to administrator and teacher needs, including providing additional training and site-based modeling of PLC structures. In Fall 2016, school and classroom leaders were trained on PLCs within a Multi-Tiered Systems of Support (MTSS), which further underscored the importance of using data to inform

instructional decision making. These trainings focused on content matters such as analyzing student work and constructing common formative assessments.

In the target district, the work of a PLC was characterized by three factors: guiding questions, resources, and products. The PLCs followed four guiding questions of 1) *What do we expect all students to learn?* 2) *How do we know if they've learned it?* 3) *How will we respond when some students do not learn?* and 4) *How will we respond when some students have already learned?* Each year since the initial implementation, PLC facilitators have been expected to follow this structure when designing and reflecting on curriculum, assessment, and instruction in their weekly meetings. They utilized their time during PLC meetings to review standards, create common lessons that include remediation and extension activities, plan for student engagement, and design scales and formative assessments to measure student learning outcomes. Prior to entering the instructional cycle, PLCs were required to set conditions for professional learning which includes establishing common language and understanding of structures.

At the center of the PLC model were assessment tools and data use as it related to the academic standards. The district both adopted and adapted student assessments to inform the PLC conversations. In addition, teachers were expected to collect quarterly data in all subject areas using common assessments developed by district, monthly data in reading through a purchased literacy tool aligned with their core curriculum, and ongoing formative and anecdotal assessments in the classroom. These data were intended to inform day-to-day instructional decisions as well as to contribute to curriculum discussions in PLCs. The district provided guidance on developing common formative assessments using a seven-step process. Teachers were asked to determine (a) what to

assess, (b) how to assess, (c) assessment plan for student learning targets, (d) timeline, (e) assessment items, (f) review protocol prior to administration, and (g) proficiency criteria for standardizing data collection. These steps established a standard structure for assessment construction across PLCs at elementary and secondary levels.

In addition to developing common formative assessments, PLCs were responsible for creating 4-factor scales to measure learning targets at lesson and unit levels. These scales were designed to inform the formative assessment of student-learning outcomes through the course of lessons and units. Students identify their levels of proficiency at the start of a lesson, and teachers continually measure student progress toward achieving the standards throughout the lesson and unit. Scales, coupled with quarterly, monthly, and ongoing common data were used to inform discussions during PLCs.

School-level administrators were provided with an additional layer of training and resources to build a culture of professional learning within and across peer groups in the target district. Specifically, they were provided with continuous support for (a) monitoring district goals in their buildings, (b) making connections between PLC work and their school success plans, (c) evaluating their role in implementation, and (d) directly engaging in the PLC work. To reinforce the school leader role, specific leader behaviors were defined for setting the vision, creating infrastructures, monitoring the health of the student and staff engagement, and identifying academic and behavioral interventions to support students who need remediation or extension. Finally, school leaders were provided with targeted training on the role of data as critical to practice.

Establishing a culture of assessment beliefs and knowledge, infused with experience, is critical to ensuring assessment practices and ultimately school reform

implemented with fidelity (Burke & Wang, 2010; Coburn & Talbert, 2006; Gallagher et al, 2008; Kerr et al., 2006). Not unlike school districts across the nation, the target school district in this study has been deeply immersed in assessment-related activities within their PLCs. It is clear the district set the conditions, given the policies extended by the state and coupled with practices established in the literature, to establish a professional learning culture for assessments among school and classroom leaders. However, it is known that school and classroom leaders often are not prepared to handle data-informed decision making, which can decrease the presence and power of assessment practices (Clifford & Mason, 2013; Ulmer, 2002). The document review illustrates the target district's commitment to assessment-related activities and PLCs, two critical components of this study. It also provides context concerning what school and classroom leaders have been provided through professional development and, based on the outcomes of the study, what additional areas may need more support.

Survey Response Rates

The *Assessment Leadership Survey* (ASLS) was administered to a comprehensive population sample of 1,216 school and classroom leaders across 79 non-charter school sites in a large Florida district. School and classroom leaders were identified by the district's human resources department based on their designation as school-level administrator (SLA) or PLC facilitator. SLAs consisted of 230 principals and assistant principals at elementary, middle, and high schools. PLC facilitators included 986 classroom teachers at elementary, middle, and high schools who currently served in as a PLC facilitator assigned to specific school sites. Of the total population sampled, 392 eligible participants responded to the survey, which reflected a 32% response rate. While

this response rate approximated the goal of 40% participation, significant missing data were evident in the responses. Only 21% (n = 81) of the respondents completed the entire survey with less than 5% missing data, which is considered negligible for missingness (Dong & Peng, 2013; Langkamp, Lehma, & Lemeshow, 2010). If cases with more than negligible missing data were excluded from this study, the response rate would have reduced to 7% in the final sample. Thus, steps were taken to remove total nonresponses and partial item nonresponses from the original sample as described below.

Of the 392 total respondents, 73 clicked on the electronic mail link, gave consent, but then abandoned the survey and did not complete any other items. These respondents were eliminated, reducing the total sample to 319 respondents. Fourteen (14) of the remaining respondents indicated their school level and role but did not respond to any items pertaining to the measured variables of assessment beliefs, experiences, knowledge, and practices. These respondents also were eliminated, reducing the sample to 305 respondents. Twenty-one (21) additional respondents were eliminated due to not reporting years of service as classroom teacher or not reporting years of service in the role of school or classroom leader. While these individuals were included in the target population based on the file provided by district's human resources department, they did not identify holding either a school or a classroom leader role included in this study; thus, these 21 respondents were excluded from the survey population. The final sample comprised 284 cases with complete responses for school role, school type, years of experience, and frequency of PLC meetings, which represented a 23% response rate. Missing data were still evident in the final sample; however, these data were handled as addressed in subsequent sections.

Descriptive Findings

Descriptives were calculated for the survey respondents in the final sample. Table 4.1 displays the number of respondents in the final sample (n = 284) according to school type and school role (i.e., 57 SLAs, 227 PLC facilitators). Forty percent (40%) of the SLAs were principals (n = 23) and 60% were assistant principals (n = 34). Due to the smaller sample size, principals and assistant principals were combined to represent SLAs in the final sample. SLAs (n = 57) reflected 25% of the target population (n = 230) while PLC facilitators (n = 227) reflected 23% of the target population (n = 986), which suggested comparable response rates to the sampling population for school roles.

Table 4.1

Final Sample Totals

School Role	Elementary *N*	Secondary *N*	Total *N*
SLA-Principal	10	13	23
SLA-Assistant Principal	11	23	34
PLC facilitator	122	105	227
Totals	146	138	284

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

School Role and Type

As shown in Table 4.2, 146 (51%) of the respondents worked in elementary schools, and 138 (49%) at the secondary level. Like SLAs, individual grade levels were collapsed into elementary and secondary level designations for school type. Since the survey responses were kept confidential, the number of unique school sites were not extracted for this study. The representation of elementary to secondary in the final sample was almost identical to the target population (i.e., 52% for elementary and 48% for secondary). Forty-two percent (42%) of the SLAs were at the elementary level (n = 21) while 58% were at the secondary (i.e., middle and high school) levels (n = 36).

Table 4.2

Comparison of the Final Sample to the Target Population

School Role	School Type	Final Sample %	Target Population %
SLA	Elementary	42	45
	Secondary	58	55
PLC facilitator	Elementary	54	59
	Secondary	46	41
Total	Elementary	51	57
	Secondary	49	43

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

Fifty-four percent (54%) of the PLC facilitators were at the elementary level (n = 122) while 46% were at the secondary (i.e., middle and high school) levels (n = 105). As shown in Table 4.2, across school roles, there were slight differences between school types for SLAs and PLC facilitators. Differences between the sample and the population were no more than 5% and likely not large enough to significantly impact sampling bias.

Years of classroom experience. All respondents were asked to report number of years serving as classroom teacher, administrator, and PLC facilitator. Table 4.3 illustrates the distribution of classroom teacher experience by school role and type. Of the 57 SLA respondents, most (44%) reported 6 to 10 years of classroom teacher experience, with 33% reporting 11 to 15 years. Alternatively, for the 227 PLC facilitators, most (40%) reported 16 or more years, with 31% reporting 11 to 15 years.

Table 4.3

Number and Percentage of Classroom Teacher Years by School Role and Type

Variable	Category	1-5 Years	6-10 Years	11-15 Years	16+ Years
Role	SLA	7 (12%)	25 (44%)	19 (33%)	6 (11%)
	PLC facilitator	17 (8%)	48 (21%)	71 (31%)	91 (40%)
Type	Elementary	8 (56%)	35 (25%)	49 (34%)	54 (37%)
	Secondary	16 (12%)	38 (28%)	41 (30%)	43 (31%)

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

The respondents were fairly evenly distributed for school type with slightly increasing numbers of 6 to 10 years, 11 to 15 years, and 16 or more years across both levels. The largest number of respondents (*n* = 96, 34%) reported 16 or more years with the second largest at 11 to 15 years (*n* = 41, 30%). These data suggest that most respondents have accumulated a multi-year history of classroom teacher experience. This is not surprising given the nature of the target population. Individuals in leadership roles were selected for this study; thus, it is reasonable to assume that they would have a history of classroom teacher experience as well.

Years of administrator experience. Table 4.4 shows the distribution of administrator experience by school role and type. Different from years of classroom teacher experience, of the 57 SLA survey respondents, most (*n* = 20, 35%) reported 1 to 5 years of administrator experience with 6 to 10 years comprising the second most years of administrator experience (*n* = 17, 30%). No PLC facilitators reported administrative experience, which matches the target survey population. Differences were observed in survey respondents for elementary compared to secondary. Most of the elementary level respondents reported either 1 to 5 years of administrator experience (*n* = 7, 29%) or 16 or more years (*n* = 9, 38%) whereas most secondary level respondents reported 1 to 5 years (*n* = 13, 39%) and 6 to 10 years (*n* = 12, 36%) experience.

Table 4.4

Number and Percentage of Administrator Years by School Role and Type

Variable	Category	1-5 Years	6-10 Years	11-15 Years	16+ Years
Role	SLA	20 (35%)	17 (30%)	8 (14%)	12 (21%)
	PLC facilitator	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Type	Elementary	7 (29%)	5 (21%)	3 (13%)	9 (38%)
	Secondary	13 (40%)	12 (36%)	5 (15%)	3 (9%)

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

Years of PLC facilitator experience. Table 4.5 shows the distribution of PLC facilitator experience by school role and type. Given the implementation of PLC initiatives over the past 4 years in the target district, survey respondents were asked to report number of years based in increments of 0 through 4 years. Of the 284 survey respondents, most (*n* = 16, 6%) reported some experience as a PLC facilitator. As expected, PLC facilitators reported at least 1 year of experience in their school role. Nearly half of the survey respondents (*n* = 92, 41%) reported serving in this role for at least 4 years while an additional quarter (*n* = 58, 26%) reported 3 years. PLC facilitators with 3 or more years in the school role represented a larger percentage of respondents than PLC facilitators with less than 3 years.

Table 4.5

Number and Percentage of PLC Facilitator Years by School Role and Type

Variable	Category	None	1 Year	2 Years	3 Years	4 Years
Role	SLA	16 (28%)	1 (2%)	5 (9%)	6 (11%)	29 (51%)
	PLC facilitator	0 (0%)	41 (18%)	36 (16%)	58 (26%)	92 (41%)
Type	Elementary	8 (6%)	26 (18%)	21 (14%)	32 (22%)	59 (40%)
	Secondary	8 (6%)	16 (12%)	20 (15%)	32 (23%)	62 (45%)

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

Interestingly, 29 (501%) of the SLAs also reported 4 years in the school role of PLC facilitator. An additional 12 (21%) SLAs reported between 1 and 3 years. Cross-tabulation analyses suggested SLAs who reported previously serving as a PLC facilitator tended to have 1 to 10 years experiences in their administrative role (*n* = 37, 65%). Several explanations may account for this outcome. SLAs may perceive themselves as PLC facilitators if they are participants in a PLC or if they are responsible for implementing the PLC model at their schools. Further, some SLAs may also serve as a

PLC facilitator at their respective schools or among their peers in other contexts. More information is needed to verify these results.

PLC meeting patterns. Respondents were asked to report on the frequency of PLC meetings as a measure of the professional learning opportunities offered at their schools as it aligns with the target district's initiatives. Table 4.6 illustrates the distribution of years of classroom teacher experience by school role and type. Of the 284 survey respondents, nearly 80% reported engaging in weekly PLC meetings. There were slight differences across school role and type, although only marginal. This pattern was consistent with the target district's initiative that by design expects weekly PLCs.

Table 4.6

Number and Percentage of PLC Meeting Frequencies by School Role and Type

Variable	Category	Daily	Weekly	Bi-Weekly	Monthly	Quarterly
Role	SLA	5 (9%)	43 (756%)	8 (14%)	1 (2%)	0 (0%)
	PLC facilitator	6 (3%)	183 (81%)	19 (9%)	12 (5%)	7 (3%)
Type	Elementary	9 (6%)	109 (75%)	21 (14%)	5 (3%)	2 (1%)
	Secondary	2 (1%)	117 (85%)	6 (4%)	8 (6%)	5 (4%)

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

Assessment Learning Experiences

Assessment learning experiences were defined as the *number of assessment-related professional learning opportunities during preservice and inservice education.* The assessment learning experiences (EXP) variable on the ASLS was measured using two discreet factors: (1) number of postsecondary courses taken during preservice (EPS) and (2) number of professional development sessions taken during inservice (EPD). The results presented below emerged from pooling mean responses across imputed datasets.

The pooled mean response for Postsecondary Courses experience (EPS) was 2.81 (*SD* = 4.72) whereas the pooled mean response for Professional Development Sessions

experience (EPD) was 8.31 (*SD* = 14.28). In the original dataset, the reported estimates of EPS experience ranged from 0 courses (*n* = 25) to 30 courses (*n* = 1). EPD ranged from 0 sessions (*n* = 48) to 150 sessions (*n* = 2). On average, respondents reported experiencing four times more inservice professional development sessions than preservice postsecondary courses in assessment. As shown in Table 4.7, independent sample t-tests yielded no significant differences for assessment learning experiences by school role or type. Across job roles and levels, respondents reported similar assessment learning experiences for postsecondary courses and professional development sessions, with more occurring during inservice experiences compared to preservice experiences.

Table 4.7

Means for Assessment Learning Experience by School Role and Type

Item	Group	*M*	*t*	*p*
Postsecondary courses (EPS)	PLC	2.99	1.21	.23
	SLA	2.09		
	Elementary	2.40	-1.20	.23
	Secondary	3.23		
Professional development sessions (EPD)	PLC	8.55	0.54	.59
	SLA	7.36		
	Elementary	6.95	-1.49	.14
	Secondary	9.76		

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

Assessment Beliefs

Assessment beliefs were defined as *attitudes that assessment practices are essential components of school-reform efforts*. The assessment beliefs (BEL) variable on the ASLS was measured using a 4-point Likert-scale that asked respondents to indicate the degree to which they agreed or disagreed with the statements provided. As shown in Table 4.8, the pooled mean for the Total BEL Score was 40.03 (*SD* = 6.37). The pooled item means ranged from 1.88 (*SD* = 0.86) for *leaders promoting effective assessment*

practices is as important as leaders promoting teaching practices to 3.41 (*SD* = 0.80) for *student context (e.g., prior knowledge, experiences, motivations, attitudes, learning styles) is as important as content when deciding HOW to teach*.

Table 4.8

Pooled Item Means and Standard Deviations for Assessment Beliefs

Item	*M*	*SD*
Q1. Data…understanding what students KNOW.	2.98	0.87
Q2. Data…understanding what students CAN DO.	2.89	0.92
Q3. Data…deciding WHAT TYPES of assessments.	2.79	0.83
Q4. Data…deciding WHICH TEACHERS are assigned.	2.62	0.89
Q5. Data…deciding about WHAT to teach.	2.90	0.87
Q6. Data…deciding about HOW to teach.	2.96	0.91
Q7. Classroom context…deciding WHAT to teach.	2.97	0.91
Q8. Classroom context…deciding HOW to teach.	3.26	0.86
Q9. Classroom context…assessing student learning.	3.12	0.86
Q10. Student context…deciding WHAT to teach.	3.11	0.93
Q11. Student context…deciding HOW to teach.	3.41	0.80
Q12. Student context…assessing student learning.	3.20	0.87
Q13. Teachers analyzing … is as important as teaching.	1.88	0.92
Q14. Leaders promoting… is as important as teaching.	1.95	0.86
Total BEL Score	40.03	6.37

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

On average, respondents exhibited the most agreement for items that pertained to the importance of context and content when making decisions about teaching and assessing. Respondents were split between agree and disagree for items that reflected student assessment data as the primary source for understanding what students know and can do as well as for making decisions about teaching. Most disagreement was for items that compared the importance of student assessment practices to teaching practices. Respondents seemed to disagree or strongly disagree with this notion compared to the other belief items. It was clear across items that while respondents tended to believe

assessment practices are important, they did not report believing they are as important as teaching practices. Respondents showed the strongest belief that both classroom and student contexts are more important than student assessment data when making teaching and assessing decisions.

By school role. As shown in Table 4.9, independent sample t-tests indicated significant differences between PLC facilitators and School-Level Administrators (SLA) for two pooled mean items on the BEL measure. SLAs (M = 2.94) reported significantly more agreement than PLC facilitators (M = 2.54) about *using student assessment data to determine which teachers should teach certain content areas*, $t(4,284) = -2.94$, $p < .001$. Alternatively, SLAs (M = 1.59) reported significantly more disagreement than PLC facilitators (M = 1.96) that *teachers analyzing multiple forms of student assessment data is as important as teaching*, $t(9,686) = 2.62$, $p = .01$. While most respondents disagreed with this item, SLAs disagreed to a greater degree. The Total BEL Score did not differ significantly by school role, $t(54,044) = -0.83$, $p = .41$, suggesting SLA and PLC facilitator groups reported similar overall belief about assessments.

Table 4.9

Pooled Independent Samples t-tests for BEL by School Role

Item	Group	*M*	*t*	*p*
Q1. Data…understanding what students KNOW.	PLC	2.95	-1.12	.26
	SLA	3.10		
Q2. Data…understanding what students CAN DO.	PLC	2.85	-1.43	.15
	SLA	3.05		
Q3. Data…deciding TYPES of assessments to use.	PLC	2.79	0.20	.84
	SLA	2.77		
Q4. Data…deciding WHICH TEACHERS assigned.	PLC	2.54	-2.94	.00
	SLA	2.94		
Q5. Data…deciding WHAT to teach.	PLC	2.87	-1.05	.29
	SLA	3.01		
Q6. Data…deciding HOW to teach.	PLC	2.92	-1.20	.23
	SLA	3.09		
Q7. Classroom context…deciding WHAT to teach.	PLC	2.96	-0.29	.77
	SLA	3.00		
Q8. Classroom context…deciding HOW to teach.	PLC	3.24	-0.70	.49
	SLA	3.33		
Q9. Classroom context…assessing student learning.	PLC	3.13	0.29	.77
	SLA	3.09		
Q10. Student context…deciding WHAT to teach.	PLC	3.10	-0.13	.90
	SLA	3.12		
Q11. Student context…deciding HOW to teach.	PLC	3.38	-1.27	.21
	SLA	3.53		
Q12. Student context…assessing student learning.	PLC	3.20	-0.13	.90
	SLA	3.21		
Q13. Teachers analyzing...as important as teaching.	PLC	1.96	2.62	.01
	SLA	1.59		
Q14. Leaders promoting…as important as teaching.	PLC	2.00	1.69	.10
	SLA	1.77		
Total BEL Score	PLC	39.89	-0.83	.41
	SLA	40.60		

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

By school type. As shown in Table 4.10, independent sample t-tests indicated significant differences between PLC facilitators and School-Level Administrators (SLA) for one pooled mean item on the BEL measure. Secondary level educators (M = 2.06) reported significantly more agreement about the importance of *leaders promoting both effective assessment and teaching practices* than elementary level educators (M = 1.85), $t(7,803) = -2.02$, $p = .04$. As mentioned, most respondents tended to disagree with this item; however, respondents at the elementary level tended to disagree more. The Total BEL Score for the assessment belief measure did not differ significantly by school type, $t(89,637) = 0.60$, $p = .55$, suggesting both elementary and secondary groups reported similar degrees of overall belief about assessments.

Assessment Practices

Assessment practices were defined as *assessment activities centered on improving student-learning outcomes such as setting an assessment vision, self-reflecting on assessment skills, identifying student-learning targets, creating formative and summative assessments matched to targets, collecting and analyzing student data at designated intervals, and adjusting instruction based on student data*. The assessment practices (PRA) variable on the ASLS was measured using a 5-point Likert-scale that asked respondents to indicate their level of engagement in assessment leadership practices.

Table 4.10

Pooled Independent Samples t-tests for BEL by School Type

Item	Group	*M*	t	*p*
Q1. Data…understanding what students KNOW.	Elementary	3.03	1.00	.32
	Secondary	2.93		
Q2. Data…understanding what students CAN DO.	Elementary	2.97	1.45	.15
	Secondary	2.81		
Q3. Data…deciding TYPES of assessments to use.	Elementary	2.84	1.12	.26
	Secondary	2.73		
Q4. Data…deciding WHICH TEACHERS assigned.	Elementary	2.65	0.52	.60
	Secondary	2.59		
Q5. Data…deciding WHAT to teach.	Elementary	2.93	0.66	.51
	Secondary	2.86		
Q6. Data…deciding HOW to teach.	Elementary	3.04	1.48	.14
	Secondary	2.87		
Q7. Classroom context…deciding WHAT to teach.	Elementary	3.05	1.59	.11
	Secondary	2.88		
Q8. Classroom context…deciding HOW to teach.	Elementary	3.28	0.49	.64
	Secondary	3.23		
Q9. Classroom context…assessing student learning.	Elementary	3.17	0.84	.40
	Secondary	3.08		
Q10. Student context…deciding WHAT to teach.	Elementary	3.06	-0.78	.44
	Secondary	3.15		
Q11. Student context…deciding HOW to teach.	Elementary	3.38	-0.71	.48
	Secondary	3.45		
Q12. Student context…assessing student learning.	Elementary	3.15	-1.00	.32
	Secondary	3.25		
Q13. Teachers analyzing...as important as teaching.	Elementary	1.83	-0.90	.37
	Secondary	1.94		
Q14. Leaders promoting…as important as teaching.	Elementary	1.85	-2.02	.04
	Secondary	2.06		
Total BEL Score	Elementary	40.23	0.60	.55
	Secondary	39.82		

As shown in Table 4.11, the pooled mean for the Total PRA Score was 66.69 (*SD* = 11.62). The pooled item means ranged from 3.08 (*SD* = 1.18) for *I provide guidance to teachers on organizing assessment data into charts and graphs* to 3.87 (*SD* = 1.16) for *I set expectations that teachers develop formative classroom assessments aligned with student-learning targets*. On average, most of the respondents reported engaging in assessment practices often (3) to frequently (4) with some relative differences for a few items. Respondents reported the least engagement in *providing guidance to teachers on writing and scoring assessment items* as well as in *organizing assessment data into charts and graphs* compared to the other assessment practices items. They reported the most engagement in *setting expectations for developing and using assessments* and *self-reflecting on their own assessment knowledge*.

Table 4.11

Pooled Item Means and Standard Deviations for Assessment Practices

Item	*M*	*SD*
Q1. school's vision for student assessments	3.24	1.21
Q2. reflective dialogue about student assessments	3.50	1.21
Q3. instructional planning based on student assessment data	3.48	1.18
Q4. reflect on my own assessment knowledge	3.74	1.26
Q5. writing assessment items	3.14	1.21
Q6. scoring assessment items	3.19	1.21
Q7. organizing data into charts and graphs	3.08	1.18
Q8. analyzing student assessment data trends	3.39	1.23
Q9. using student assessment data to adjust instruction	3.49	1.26
Q10. using student assessment data to change assessment items	3.12	1.16
Q11. identify student-learning targets	3.86	1.26
Q12. develop formative classroom assessments	3.87	1.16
Q13. design summative classroom assessments	3.74	1.21
Q14. collect multiple forms of student-assessment data	3.71	1.22
Q15. analyze multiple forms of student-assessment data	3.58	1.25
Q16. modify curriculum based on outcomes	3.51	1.22
Q17. adjust instruction based on outcomes	3.75	1.16
Q18. select assessment tools based on outcomes	3.44	1.19
Total PRA	66.69	11.62

By school role. As shown in Table 4.12, independent sample t-tests indicated significant differences between PLC facilitators and School-Level Administrators (SLA) for two items on the PRA measure. SLAs (M = 3.40) reported *providing guidance to teachers on organizing assessment data into charts and graphs* more often than PLC facilitators (M = 3.00), $t(1,307) = -2.05$, $p = .04$. Similarly, SLAs (M = 3.69) reported *providing guidance to teachers on analyzing student assessment data* more often than PLC facilitators (M = 3.31), $t(1,447) = -1.90$, $p = 0.05$. The remaining items on the practices measure did not evidence significant differences between school roles. The Total PRA Score for the assessment practices measure did not differ significantly by school role, $t(89,671) = -1.62$, $p = .21$, suggesting both PLC facilitator and SLA groups reported similar levels of engagement in assessment practices.

Table 4.12

Pooled Independent Samples t-tests for PRA by School Role

Item	Group	*M*	t	*p*
Q1. school's vision for student assessments	PLC	3.17	-1.81	.07
	SLA	3.54		
Q2. reflective dialogue about student assessments	PLC	3.45	-1.27	.21
	SLA	3.70		
Q3. instructional planning based on student data	PLC	3.47	-0.17	.87
	SLA	3.51		
Q4. reflect on my own assessment knowledge	PLC	3.76	0.49	.63
	SLA	3.66		
Q5. writing assessment items	PLC	3.21	1.85	.07
	SLA	2.84		
Q6. scoring assessment items	PLC	3.21	0.384	.70
	SLA	3.13		
Q7. organizing data into charts and graphs	PLC	3.00	-2.05	.04
	SLA	3.40		
Q8. analyzing student data trends	PLC	3.31	-1.90	.05
	SLA	3.69		
Q9. using student data to adjust instruction	PLC	3.43	-1.33	.19
	SLA	3.72		
Q10. using student data to change assessment items	PLC	3.13	0.29	.77
	SLA	3.08		
Q11. identify student-learning targets	PLC	3.82	-0.93	.34
	SLA	4.02		
Q12. develop formative classroom assessments	PLC	3.86	-0.35	.73
	SLA	3.92		
Q13. design summative classroom assessments	PLC	3.69	-1.34	.18
	SLA	3.95		
Q14. collect multiple forms of student data	PLC	3.66	-1.36	.17
	SLA	3.93		
Q15. analyze multiple forms of student data	PLC	3.50	-1.92	.06
	SLA	3.89		
Q16. modify curriculum based on outcomes	PLC	3.46	1.18	.24
	SLA	3.70		
Q17. adjust instruction based on outcomes	PLC	3.70	-1.28	.20
	SLA	3.95		
Q18. select assessment tools based on outcomes	PLC	3.41	-0.93	.35
	SLA	3.59		
Total PRA Score	PLC	66.17	-1.62	.12
	SLA	68.72		

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

By school type. As shown in Table 4.13, independent sample t-tests indicated significant differences between elementary level and secondary level groups for three items on the PRA measure. Elementary level educators (M = 3.41) reported discussing with teachers the school's vision for student assessments more often than secondary level educators (M = 3.41), t(1,078) = 2.11, p = .04. In addition, elementary level educators (M = 3.68) reported using student data to plan instruction more often than secondary level educators (M = 3.27), t(1,398) = 2.69, p = 0.00. Like SLAs, elementary level educators (M = 3.23) reported spending relatively more time organizing assessment data into charts and graphs compared to secondary level educators (M = 2.93), t(1,091) = 1.89, p = .05. The Total PRA Score for the assessment practices measure differed significantly by school type, t(72,189) = 2.61, p = .01, which suggested that elementary level educators (M = 68.28) tend to engage in assessment practices as defined by this scale more often than secondary level educators (M = 65.00).

Table 4.13

Pooled Independent Samples t-tests for PRA by School Type

Item	Group	*M*	t	*p*
Q1. school's vision for student assessments	Elementary	3.41	2.11	.04
	Secondary	3.07		
Q2. reflective dialogue about student assessments	Elementary	3.63	1.75	.08
	Secondary	3.36		
Q3. instructional planning based on student data	Elementary	3.68	2.69	.00
	Secondary	3.27		
Q4. reflect on my own assessment knowledge	Elementary	3.82	1.02	.31
	Secondary	3.65		
Q5. writing assessment items	Elementary	3.17	0.38	.71
	Secondary	3.11		
Q6. scoring assessment items	Elementary	3.28	1.08	.28
	Secondary	3.11		
Q7. organizing data into charts and graphs	Elementary	3.23	1.89	.05
	Secondary	2.93		
Q8. analyzing student data trends	Elementary	3.49	1.25	.21
	Secondary	3.28		
Q9. using student data to adjust instruction	Elementary	3.62	1.66	.10
	Secondary	3.34		
Q10. using student data to change assessment items	Elementary	3.20	1.04	.30
	Secondary	3.04		
Q11. identify student-learning targets	Elementary	3.96	1.15	.25
	Secondary	3.77		
Q12. develop formative classroom assessments	Elementary	3.94	1.00	.32
	Secondary	3.79		
Q13. design summative classroom assessments	Elementary	3.76	0.26	.80
	Secondary	3.72		
Q14. collect multiple forms of student data	Elementary	3.79	1.00	.32
	Secondary	3.63		
Q15. analyze multiple forms of student data	Elementary	3.65	0.94	.35
	Secondary	3.49		
Q16. modify curriculum based on outcomes	Elementary	3.50	-0.12	.91
	Secondary	3.52		
Q17. adjust instruction based on outcomes	Elementary	3.84	1.07	.29
	Secondary	3.67		
Q18. select assessment tools based on outcomes	Elementary	3.52	0.93	.35
	Secondary	3.36		
Total PRA Score	Elementary	68.28	2.61	.01
	Secondary	65.00		

Assessment Knowledge

Assessment knowledge was defined as *the competencies required to be assessment literate.* Assessment knowledge (ALI) was measured across seven standards using five scenarios with seven questions each where respondents had to select the correct answer select from four multiple choice responses. The seven standards are (a) choosing appropriate assessment methods; (b) developing appropriate assessment methods; (c) administering, scoring, and interpreting assessment results; (d) using assessment results to make decisions; (e) developing valid assessment procedures; (f) communicating assessment results; and (g) recognizing unethical or illegal practices. Pooled means and standard deviations are reported grouped by corresponding items.

Table 4.14

Pooled Item Means and Standard Deviations for Assessment Knowledge

Item	*M*	*SD*
S1. Choosing Appropriate Assessment Methods (Q1, Q8, Q15, Q22, Q29)	2.96	0.99
S2. Developing Appropriate Assessment Methods (Q2, Q9, Q16, Q23, Q30)	2.19	0.90
S3. Administering, Scoring, and Interpreting Results (Q3, Q10, Q17, Q24, Q31)	2.94	1.07
S4. Using Assessment Results to Make Decisions (Q4, Q11, Q18, Q25, Q32)	3.09	1.09
S5. Developing Valid Grading Procedures (Q5, Q12, Q19, Q26, Q33)	3.50	1.06
S6. Communicating Assessment Results (Q6, Q13, Q20, Q27, Q34)	3.36	1.10
S7. Recognizing Unethical or Illegal Practices (Q7, Q14, Q21, Q28, Q35)	3.66	1.02
Total ALI Score	21.88	3.28

As shown in Table 4.14, the pooled mean for the Total ALI Score was 21.88 (*SD* = 3.28) out of 35 total possible, which translated to an average score of 63% correct. In the original dataset (n = 78), the total correct ranged from 10 (29%) to 31 (89%). Across

ALI standards, mean scores were highest for Developing Valid Grading Procedures (*M* = 3.50, *SD* = 1.06) and Recognizing Unethical or Illegal Practices (*M* = 3.66, *SD* = 1.02) compared to Developing Appropriate Assessment Methods (*M* = 2.19, *SD* = 0.90).

By school role. As shown in Table 4.15, independent sample t-tests indicated significant differences between PLC facilitators and School-Level Administrators (SLA) for the Total ALI Score only. SLAs (*M* = 22.77) scored slightly higher than PLC facilitators (*M* = 21.66), *t*(1,088) = -2.19, *p* = .03. On average, SLAs obtained a total score of 65% while PLC facilitators scored a total of 62%, both groups below average. No significant differences among school roles were evident across the standards.

Table 4.15

Pooled Independent Samples t-tests for ALI by School Role

Item	Group	*M*	t	*p*
S1. Choosing Appropriate Assessment Methods (Q1, Q8, Q15, Q22, Q29)	PLC	2.93	-0.71	.48
	SLA	3.06		
S2. Developing Appropriate Assessment Methods (Q2, Q9, Q16, Q23, Q30)	PLC	2.18	-0.15	.88
	SLA	2.21		
S3. Administering, Scoring, and Interpreting Results (Q3, Q10, Q17, Q24, Q31)	PLC	2.88	-1.54	.13
	SLA	3.17		
S4. Using Assessment Results to Make Decisions (Q4, Q11, Q18, Q25, Q32)	PLC	3.06	-0.67	.50
	SLA	3.20		
S5. Developing Valid Grading Procedures (Q5, Q12, Q19, Q26, Q33)	PLC	3.45	-1.27	.20
	SLA	3.71		
S6. Communicating Assessment Results (Q6, Q13, Q20, Q27, Q34)	PLC	3.32	-1.04	.30
	SLA	3.53		
S7. Recognizing Unethical or Illegal Practices (Q7, Q14, Q21, Q28, Q35)	PLC	3.63	-0.80	.43
	SLA	3.77		
Total ALI	PLC	21.66	-2.19	.03
	SLA	22.77		

Note. SLA=School-Level Administrator, PLC=Professional Learning Community.

By school type. As shown in Table 4.16, no significant differences between elementary and secondary levels were observed for either the total score or the standards

for ALI. On average, elementary level educators scored 62% and secondary level educators scored 63%, suggesting similar performance on ALI despite the level.

Table 4.16

Pooled Item Independent Samples t-tests for ALI by School Type

Item	Group	*M*	t	*p*
S1. Choosing Appropriate Assessment Methods (Q1, Q8, Q15, Q22, Q29)	Elementary	2.95	-0.02	.99
	Secondary	2.96		
S2. Developing Appropriate Assessment Methods (Q2, Q9, Q16, Q23, Q30)	Elementary	2.23	0.66	.51
	Secondary	2.14		
S3. Administering, Scoring, and Interpreting Results (Q3, Q10, Q17, Q24, Q31)	Elementary	2.83	-1.45	.15
	Secondary	3.05		
S4. Using Assessment Results to Make Decisions (Q4, Q11, Q18, Q25, Q32)	Elementary	3.06	-0.37	.71
	Secondary	3.12		
S5. Developing Valid Grading Procedures (Q5, Q12, Q19, Q26, Q33)	Elementary	3.48	-0.26	.79
	Secondary	3.52		
S6. Communicating Assessment Results (Q6, Q13, Q20, Q27, Q34)	Elementary	3.33	-0.43	.67
	Secondary	3.40		
S7. Recognizing Unethical or Illegal Practices (Q7, Q14, Q21, Q28, Q35)	Elementary	3.61	-0.59	.56
	Secondary	3.70		
Total ALI Score	Elementary	21.75	-0.66	.51
	Secondary	22.02		

Demographics

Demographic data were collected in Section 5 of survey for descriptive purposes only. As explained in previous sections, increasing missing data was observed as the survey progressed. Since demographic information was collected in the final stage of the survey, there were significant missing values. Thus, observed data were compared to the population data when possible. Note that due to the negligible reports of ethnicities other than white, categories were collapsed and ethnicity was reported as white or other.

Table 4.17

Demographics

Variable	% Missing	Selection	% Observed	% Population
Gender	60	Male	19	18
		Female	81	82
Ethnicity	60	White	91	90
		Other	8	10
Education	59	Bachelors	38	NA
		Masters	60	NA
		Specialist	2	NA
		Doctorate	0.0	NA

Variable	% Missing	Sample $\bar{x}$	Population μ
Age	61	47	NA

As shown in Table 4.17, differences between observed gender and ethnicity in the original sample and the target population are minimal, despite missing data. In all, 81% of the sample were female. and 91% were White; these data are consistent with the population. Most respondents (60%) reported completing a master's degree as their highest-level education obtained, whereas 38% reported earning a baccalaureate degree and 1.7% earned a specialist degree. No respondents indicated holding a doctoral degree. Even though comparison data were not available, education levels were not surprising given the target population of school leaders. In the state of Florida, administrators are required to obtain a master's degree, and many teachers in leadership roles seek graduate degrees. Last, the average age reported was 47 years, which also was not surprising given that most respondents reported working 11 or more years in education.

In addition to providing demographic data, respondents were asked to verify whether they believe the assessment leadership definition was accurate and, if not, how they would recommend changing the definition. No specific procedure was applied as this item was intended to inform future research in this area. Of the 68 responses, 53 said

they agreed with the assessment leadership definition, which represents 78% of the respondents. Most of the respondents who completed this question acknowledged agreement with the four components: (a) establish a vision that sets clear and appropriate expectations for student assessment systems, (b) lead data discussions, (c) foster assessment literacy in teachers through ongoing collaborative learning experiences, and (d) self-reflect on personal assessment practices. Three respondents recommended adding a fifth element that captures creating grading systems that support students mastering standards content. Of the respondents that did not agree, most indicated that they did not have the time to carry out these components. They also reported perceptions that while it is their responsibility as PLC facilitators to follow district expectations, they did not feel adequately prepared to accomplish all components of assessment leadership.

Factor and Classical Item Analyses Findings

Prior to modeling the research questions, a series of factor analyses and classical item analyses were conducted with the original dataset, as well as with the 100 imputed datasets, to determine the simplest factor structures for assessment beliefs (BEL), practices (PRA), and knowledge (ALI). The outcomes were used to determine the most reliable factor structure for the model measures, given this dataset of service respondents.

Confirmatory Factor Analysis

First, confirmatory factor analysis (CFA) with maximum likelihood parameter (MLR) estimator was run for the assessment beliefs (BEL) measure assuming a one-factor model. Fit indices were assessed for CFI/TLI > .90, RMSEA< .06 and SRMR < .08 (Hu & Bentler, 2009). All four fit indices suggested poor fit for the one-factor model. CFA with MLR estimator was run again assuming a bi-factor model, given the grouping

of the 14 BEL items into four testlets in the survey. As shown in Table 4.18, the bi-factor model resulted in improved fit compared to the one-factor model; however, the bifactor model was still considered marginal fit compared to the recommended thresholds. Additional CFAs were run for BEL forcing two, three, and four factor solutions. The models were cross-evaluated for *good*, *marginal*, or *poor* fit (Brown, 2015). None of the factor solutions were determined as having "good" fit and only the standardized root mean-square (SRMR) suggested "marginal" fit for the multi-factor solutions. Given the "poor" to "marginal" fit of the CFAs, further model revisions were not conducted for BEL. Instead, dimensionality of the BEL measure was explored using Exploratory Factor Analysis (EFA) in IBM®SSPS® for each of the 100 imputed datasets.

Table 4.18

CFA Fit Indices by Solution for Assessment Beliefs (BEL)

Solution	CFI	TLI	RMSEA (CI)	SRMR
1-Factor Model	.636	.570	.135 (.134-.136)	.106
Bifactor Model	.876	.821	.086 (.085-.089)	.075
2-Factor Model	.791	.750	.103 (.102-.104)	.076
3-Factor Model	.825	.785	.095 (.094-.097)	.071
4-Factor Model	.858	.818	.087 (.085-.089)	.068

Note. CFI=comparative fit index; TLI=Tucker-Lewis index; RMSEA=root mean-square error of approximation; SRMR=standardized root mean-square.

Next, CFA with MLR estimator was run for the assessment practices (PRA) measure assuming a one-factor model. Like BEL, all four fit indices suggested poor fit. The CFA bifactor model was attempted for the 18 PRA items, which were organized into three distinct testlets in the survey. As shown in Table 4.18, the bifactor model also resulted in somewhat improved fit compared to the one-factor model; however, the bifactor model was still considered marginal. Like BEL, CFAs were run forcing factor solutions for two and three factor models. Given the "poor" to "marginal" fit across

models, further model revisions were not conducted for PRA. Instead, like BEL, dimensionality of the PRA measure was explored by conducting Exploratory Factor Analysis (EFA) in IBM®SSPS® for each of the 100 imputed datasets.

Table 4.19

CFA Fit Indices by Solution for Assessment Practices (PRA)

Solution	CFI	TLI	RMSEA (CI)	SRMR
1-Factor Model	.740	.705	.099 (.098-.100)	.083
Bifactor Model	.827	.772	.088 (.087-.089)	.070
2-Factor Model	.748	.712	.098 (.097-.099)	.082
3-Factor Model	.776	.740	.093 (.092-.094)	.079

Note. CFI=comparative fit index; TLI=Tucker-Lewis index; RMSEA=root mean-square error of approximation; SRMR=standardized root mean-square.

Last, CFA with MLR estimators was run for the assessment knowledge (ALI) measure assuming a one-factor, five-factor, and seven-factor model, based on established factor structures in the literature. As shown in Table 4.20, all four fit indices suggested poor fit across models. Like BEL and PRA, the dimensionality of the ALI measure was further explored by conducting Exploratory Factor Analysis (EFA) in IBM®SSPS® for each of the 100 imputed datasets.

Table 4.20

CFA Fit Indices by Solution for Assessment Knowledge (ALI)

Solution	CFI	TLI	RMSEA (CI)	SRMR
1-Factor Model	.283	.238	.075 (.074-.076)	.090
Bifactor Model	.669	.624	.053 (.052-.054)	.071
5-Factor Model	.521	.482	.062 (.061-.063)	.082
7-Factor Model	.339	.270	.075 (.075-.075)	.090

Note. CFI=comparative fit index; TLI=Tucker-Lewis index; RMSEA=root mean-square error of approximation; SRMR=standardized root mean-square.

Exploratory Factor Analysis

EFA using PAF and Direct Oblimin rotation and eigenvalues > 1.0 was run on the original as well as each of the 100 imputed datasets for the BEL measure. Prior to analysis, correlation matrices and measures of sampling adequacy were reviewed to determine suitability for factor analysis. Trends across correlation matrices showed less than ~10% significance levels above $p < .05$ with many sizable correlation coefficients over .30 (~70%) and no items above .90. Overall, BEL items appeared well correlated with little evidence of singularity. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity were averaged across imputed datasets and compared to the original dataset. KMO averaged .84 for the imputed datasets, which, like the original dataset (KMO = .82), fell above the acceptable threshold of KMO < .50 and within the *great* range of KMO < .80 (Hutcheson & Sofroniou, 1999; Kaiser, 1974). Bartlett's Tests of Sphericity were significant for both the imputed (χ^2 [91] = 1,786.77, $p< .05$) and original (χ^2 [91] = 1,649.66, $p < .05$) datasets. These results suggested BEL items were adequate for sample size and appropriate for factor analysis.

The factors that accounted for the most variance in the BEL measure were identified based on individual scree plot review for each of the imputed datasets. Across imputed datasets, the number of factors ranged from two factors (12%) to four factors (12%) with most datasets suggesting three factors (76%). Figure 4.1 illustrates the scree plot analysis for the original dataset compared to the averaged imputed datasets. While the majority of datasets with eigenvalues > 1.0 suggested a three-factor model, the scree plot for the averaged imputed datasets indicated the greatest amount of variance was explained by the first factor (38%) with an eigenvalue of 5.18 and a ratio to the second

eigenvalue of 3:1. Based on the trends across datasets, coupled with the findings from the CFA, EFA was rerun forcing a three-factor model. The factor and pattern matrices were reviewed across imputations, but did not support a consistent, discernible pattern. Given the ratio of the first eigenvalue to the second eigenvalue (Gorsuch, 1988), coupled with the marginal fit from the CFA, it was determined to treat the BEL measure as unidimensional. BEL was rerun forcing a one-factor model and analyzed for loadings.

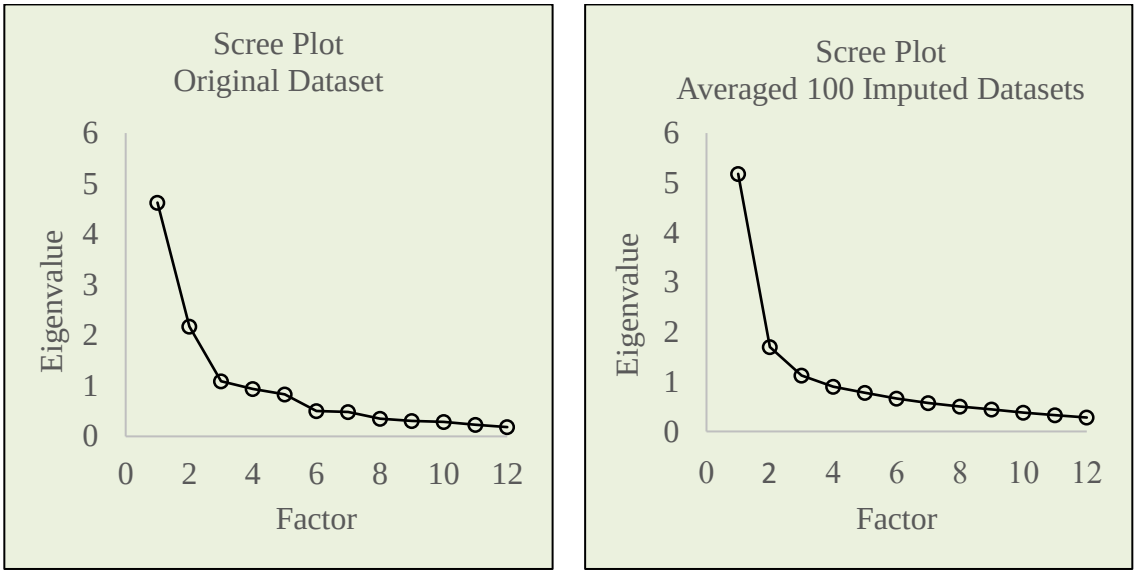

Figure 4.1. Scree plot analysis for assessment beliefs (BEL).

Several steps were taken to determine the simplest factor structure for BEL assuming a unidimensional scale. First, communalities (C) were reviewed to identify any items with shared variances below .50. As displayed in Table 4.21, all items were moderately correlated above the minimum acceptable shared variance of .3 with an average of .50. It should be noted that questions two through six fell below the desired threshold of .50 with question four below .40, which suggested this item had the lowest common variance with the other items. Next, the factor matrix (F) was examined to determine the size of the factor loadings. While all averaged factor loadings were above

the .40 threshold, 37% of the individual imputed datasets contained factor loadings below .40 for question 4, which was consistent with the least shared variance identified in the communalities. The next most frequent items below .40 comprised 6% of the datasets.

Table 4.21

Full One-Factor Solution for Assessment Beliefs (BEL)

Item	h^2	λ	CITC
Q1. Data…understanding what students KNOW.	.54	.66	.53
Q2. Data…understanding what students CAN DO.	.48	.58	.49
Q3. Data…deciding WHAT TYPES of assessments.	.40	.58	.49
Q4. Data…deciding WHICH TEACHERS are assigned.	.33	.42	.31
Q5. Data…deciding about WHAT to teach.	.45	.58	.49
Q6. Data…deciding about HOW to teach.	.49	.62	.46
Q7. Classroom context…deciding WHAT to teach.	.58	.60	.55
Q8. Classroom context…deciding HOW to teach.	.53	.61	.56
Q9. Classroom context…assessing student learning.	.58	.58	.55
Q10. Student context…deciding WHAT to teach.	.51	.56	.55
Q11. Student context…deciding HOW to teach.	.53	.59	.53
Q12. Student context…assessing student learning.	.55	.62	.60
Q13. Teachers analyzing…is as important as teaching.	.46	-.51	-.37
Q14. Leaders promoting…is as important as teaching.	.52	-.56	-.42

Note. h^2 = communality, λ=factor loading, CITC=corrected item-total correlation.

Coefficient alpha was calculated to evaluate the internal consistency of the items. Reliability statistics were computed and averaged for the imputed datasets and compared to the original dataset. Across imputed datasets, coefficient alpha ranged from α = .67 to α = .82 with an average of α = .74. The corrected item-total correlations (CITC) were consistently above the desired threshold of .40, except for question 4, which was at .30. Trends across selection criteria indicated that question 4 was not favorable for the BEL measure. Thus, question 4, *I believe student assessment data are the primary source for making decisions about WHICH TEACHERS are assigned to teach certain content areas,*

was removed from the instrument. Reliability statistics for questions 13 and 14 suggested lower internal consistency and improved coefficient alpha if deleted.

Table 4.22

Revised One-Factor Solutions for Assessment Beliefs (BEL)

	λ	
Item	(-) Q4	(-) Q4, Q13, Q14
Q1. Data…understanding what students KNOW.	.64	.61
Q2. Data…understanding what students CAN DO.	.56	.54
Q3. Data…deciding WHAT TYPES of assessments.	.57	.56
Q4. Data…deciding WHICH TEACHERS are assigned.	-	-
Q5. Data…deciding about WHAT to teach.	.56	.54
Q6. Data…deciding about HOW to teach.	.60	.54
Q7. Classroom context…deciding WHAT to teach.	.61	.64
Q8. Classroom context…deciding HOW to teach.	.63	.65
Q9. Classroom context…assessing student learning.	.60	.64
Q10. Student context…deciding WHAT to teach.	.58	.61
Q11. Student context…deciding HOW to teach.	.60	.62
Q12. Student context…assessing student learning.	.63	.66
Q13. Teachers analyzing…is as important as teaching.	-.49	-
Q14. Leaders promoting…is as important as teaching.	-.54	-

Note. λ=factor loading.

Finally, EFA using PAF with Direct Oblimin rotation was conducted on the 100 imputed datasets without question 4 and again without questions 13 and 14. As shown in Table 4.22, the averaged BEL items across imputed datasets had factor loadings at or above the desired threshold of .50 except for question 13 in the (-) Q4 factor solution. Coefficient alpha for the (-) Q4 factor solution ranged from $\alpha = .66$ to $\alpha = .81$ with an average of $\alpha = .74$, which was similar to the reliability statistics when all items were included in the model. Internal consistency in these ranges are considered poor to acceptable (Brown, 2015). Thus, EFA was rerun without questions 4, 13 and 14. Loadings for the (-) Q4, Q13, Q14 one-factor solution improved to more desirable thresholds of .50 and .60 with coefficient alpha ranging from $\alpha = .83$ to $\alpha = .91$ with an

average of α = .86. The improved loadings coupled with communalities and reliabilities suggested that (-) Q4, Q13, Q14 was the best solution for BEL.

Like BEL, EFA using PAF and Direct Oblimin rotation and eigenvalues > 1.0 was run on the original as well as each of the 100 imputed datasets for the PRA measure. Trends across correlation matrices showed less than ~10% significance levels above $p < .05$ with many sizable correlation coefficients over .30 (~65%) and no items above .90. PRA items appeared well correlated with little evidence of singularity. Moreover, KMO averaged .88 for the imputed datasets, which, like the original dataset (KMO = .94), fell above the acceptable threshold of KMO < .50 and within the *great* range of KMO < .80. Bartlett's Tests of Sphericity were significant for both the imputed (χ^2 [153] = 2,34.51, $p < .05$) and original (χ^2 [153] = 2,281.51, $p < .05$) datasets. The results suggested that PRA items were adequate for sample size and appropriate for factor analysis.

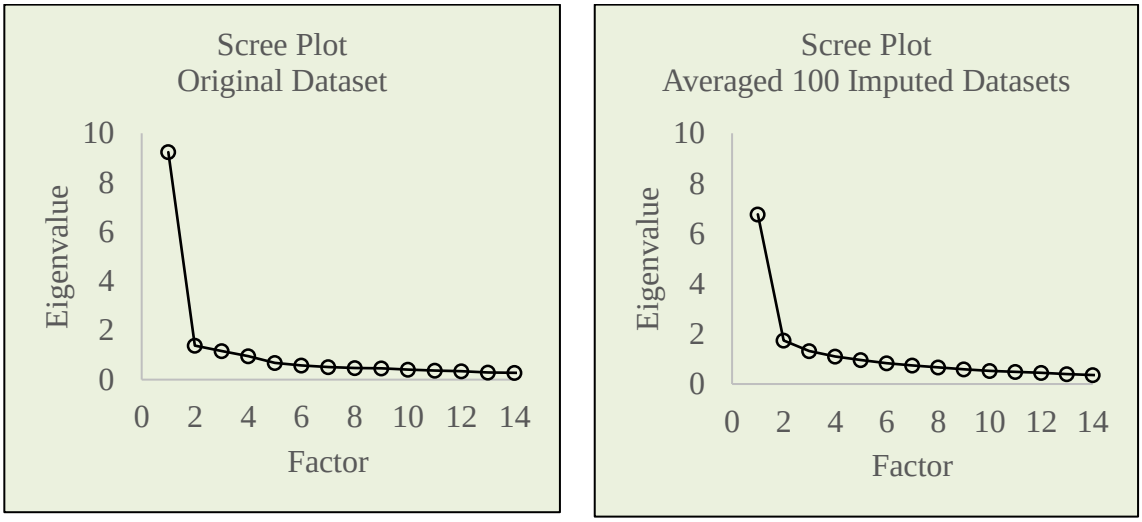

Figure 4.2. Scree plot analysis for assessment practices (PRA).

The factors that accounted for the most variance in the PRA measure were identified based on individual scree plots review for each of the imputed datasets. Across imputed datasets, the number of factors ranged from three factors (10%) to five factors

(26%) with most datasets suggesting three factors (64%). Figure 4.2 illustrates the scree plot analysis for the original datasets compared to the averaged imputed datasets. Like BEL, while the majority of datasets with eigenvalues > 1.0 supported a four-factor model, the scree plot for the averaged imputed datasets suggested the greatest amount of variance was explained by the first factor (38%) with an eigenvalue of 6.77 and a ratio to the second eigenvalue of 4:1. Based on the trends across datasets, coupled with the findings from the CFA, EFA was rerun forcing a four-factor model. Like BEL, the factor and pattern matrices were reviewed across imputations, but did not support a consistent discernible pattern. Thus, it was determined to treat the PRA measure as unidimensional (Gorsuch, 1988). PRA was rerun forcing a one-factor model and analyzed for loadings.

Table 4.23

Full One-Factor Model for Assessment Practices (PRA)

Item	h^2	λ	CITC
Q1. school's vision for student assessments	.50	.50	.48
Q2. reflective dialogue about student assessments	.66	.67	.62
Q3. instructional planning based on student data	.54	.54	.52
Q4. reflect on my own assessment knowledge	.52	.62	.57
Q5. writing assessment items	.45	.42	.41
Q6. scoring assessment items	.58	.52	.51
Q7. organizing data into charts and graphs	.40	.38	.37
Q8. analyzing student assessment data trends	.60	.57	.54
Q9. using student assessment data to adjust instruction	.70	.65	.61
Q10. using student data to change assessment items	.54	.49	.49
Q11. identify student-learning targets	.60	.66	.61
Q12. develop formative classroom assessments	.69	.62	.59
Q13. design summative classroom assessments	.63	.56	.56
Q14. collect multiple forms of student-assessment data	.67	.60	.57
Q15. analyze multiple forms of student-assessment data	.67	.64	.60
Q16. modify curriculum based on outcomes	.55	.58	.55
Q17. adjust instruction based on outcomes	.71	.66	.63
Q18. select assessment tools based on outcomes	.60	.56	.53

Note. h^2 = communality, λ=factor loading, CITC=corrected item-total correlation.

The same approach used for BEL was applied to PRA to determine the simplest factor structure assuming a unidimensional scale. As shown in Table 4.23, all PRA items were moderately correlated well above the minimum acceptable shared variance of .30 with an average of .59. It should be noted that questions 5 and 7 fell below the desired threshold of .50, but not below .40, which suggests these items shared the least common variance with the other items, which is questionable given that 40% of the variance is due to measurement error. Next, the factor matrix (F) was examined to determine the size of the factor loadings. Of the three item questions with factor loadings below the .50 threshold, 53% of the individual imputed datasets were below .40 for question 7, whereas 17% and 12% were below .40 for questions 5 and 10, respectively.

Table 4.24

Revised One-Factor Solution (-) Q7 for Assessment Practices (PRA)

Item	λ
Q1. school's vision for student assessments	0.62
Q2. reflective dialogue about student assessments	0.76
Q3. instructional planning based on student data	0.72
Q4. reflect on my own assessment knowledge	0.66
Q5. writing assessment items	0.58
Q6. scoring assessment items	0.69
Q7. organizing data into charts and graphs	-
Q8. analyzing student assessment data trends	0.67
Q9. using student assessment data to adjust instruction	0.79
Q10. using student data to change assessment items	0.60
Q11. identify student-learning targets	0.73
Q12. develop formative classroom assessments	0.75
Q13. design summative classroom assessments	0.70
Q14. collect multiple forms of student-assessment data	0.73
Q15. analyze multiple forms of student-assessment data	0.75
Q16. modify curriculum based on outcomes	0.69
Q17. adjust instruction based on outcomes	0.84
Q18. select assessment tools based on outcomes	0.75

Note. λ=factor loading.

Coefficient alpha was calculated to evaluate the internal consistency of the PRA items. Reliability statistics were computed and averaged for the imputed datasets and compared to the original dataset. Across imputed datasets, coefficient alpha ranged from $\alpha = .88$ to $\alpha = .93$ with an average of $\alpha = .90$, which suggested excellent internal consistency. The corrected item-totals (CITC) were above the desired threshold of .40, except for question 7. Trends across selection criteria indicated that question 7 was not favorable for the PRA measure. Thus, question 7, *I provide guidance to teachers on organizing assessment data into charts and graphs* was removed from the model. Finally, EFA using PAF with Direct Oblimin rotation was conducted again on the 100 imputed datasets without question 7. As shown in Table 4.24, the averaged PRA items across imputed datasets had factor loadings at or above the desired threshold of .50. Coefficient alpha for the (-) Q7 factor solution remained above $\alpha = .90$, which suggests internal consistency maintained within excellent range despite item deletion. The improved loadings coupled with strong reliabilities suggests (-) Q7 is the best solution.

Like BEL and PRA, EFA using PAF and Direct Oblimin rotation and eigenvalues > 1.0 was run on the original as well as each of the 100 imputed datasets for the ALI measure. Trends across correlation matrices were significantly less favorable for ALI than for BEL and PRA. Most correlation coefficients were below .30 (~95%) with over half the significance levels were above $p < .05$ (~59%). Similarly, KMO averaged .66 for the imputed datasets, which, unlike the original dataset (KMO = .45), fell just above the acceptable threshold of KMO < .50 (Hutcheson & Sofroniou, 1999; Kaiser, 1974). Bartlett's Tests of Sphericity for ALI were significant for both the imputed ($\chi^2[595] =$

1,831.93, $p < .05$) and original ($\chi^2[595] = 692.57$, $p < .05$) datasets. Results suggested that the ALI measure was only marginally adequate for sample size and factor analysis

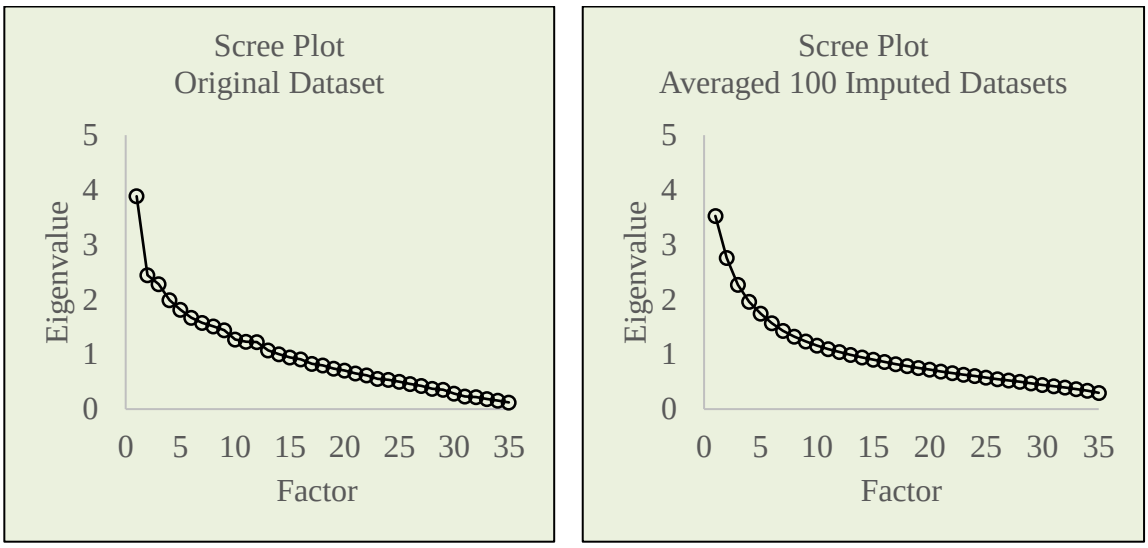

Figure 4.3. Scree plot analysis for assessment knowledge (ALI).

Even so, scree plots were reviewed to determine how many factors accounted for the most variance in the ALI measure. Across imputed datasets, the number of factors ranged from 11 factors (22%) to 14 factors (1%) with most datasets suggesting 12 factors (48%) or 13 factors (30%). Figure 4.3 illustrates the scree plot analysis for the original datasets compared to the averaged imputed datasets. Only 10% of the variance was explained by the first factor, with up to 40% explained by the first six factors. Communalities suggested the average shared variance among items was .29, which suggests more of the variance is explained by other factors not measured in the scale. Given the EFA findings, coupled with the relative poor fit observed in CFA, it was determined to conduct classical item analysis to determine which items to retain.

Classical Item Analysis

IBM®SSPS® was employed to conduct classical item analysis using reliability statistics and item-total correlations. The results were averaged across the 100 imputed

datasets. Across 100 imputed datasets, coefficient alpha for the 35-item ALI ranged from $\alpha = .38$ to $\alpha = .61$ with an average of $\alpha = .49$, which suggested poor internal consistency. As displayed in Table 4.25, the corrected item-total correlations (CITCs) for the 35-item model averaged .12, which was a similar trend as the low shared variances observed in the Communalities from the EFA. Eight (8) items were removed based on the coefficient alpha if deleted values. With the removal of these items, reliability improved for the 27 remaining items to an average of $\alpha = .63$ with an average CITC of .19. None of the values for coefficient alpha if deleted suggested improved reliability with additional deletions. Instead, the corrected item-total correlation below .20 was used to eliminate 2 other items. Unfortunately, reliability did not improve with the additional item deletions ($\alpha = .49$). Thus, the 27-item measure was retained as the final structure.

As shown in Table 4.25, the poorest performing items occurred in ALI Standards 1, 2, and 3. These standards contained items related to choosing and developing appropriate assessment methods and administering, scoring, and interpreting results. Although items were removed to improve the internal reliability of the measure, it should be noted that these items were more problematic for respondents on average, as evidenced by the lower pooled mean scores discussed in previous sections. Standard related to communicating assessment results, using assessments to make decisions, developing valid grading practices, and recognizing ethical practices performed relatively better than the other items, both for the factor structure and based on the pooled means.

Table 4.25

Corrected Item-Total Correlations for Assessment Knowledge (ALI)

Item	35-items	27-items	24-items
S1. Choosing Appropriate Assessment Methods			
Q1 …type of assessment that best answers	-.06	-	-
Q8 …most appropriate assessment type	.25	.21	.20
Q15 …what can you conclude about the decision	.22	-	-
Q22 …method is the best to answer the question	-.02	.20	.21
Q29 …assessment is the best to meet the needs	.22	.20	.20
S2. Developing Appropriate Assessment Methods			
Q2 …grading accurately	.25	.18	-
Q9 …choosing appropriate type of assessment	.22	-	-
Q16 …developing quality multiple choice tests	-.02	-	-
Q23 …recommending items for story-based tests	.22	.20	.10
Q30 …discarding or revising test items	.25	-	-
S3. Administering, Scoring, and Interpreting Results			
Q3 …administering standardized math tests	.22	.20	-
Q10 …interpreting percentile range	-.02	-	-
Q17 …interpreting scores on 100-percent scale	.22	.19	.28
Q24 …comparing means and standard deviations	.25	.20	.07
Q31 …scoring restricted response essays	.19	.19	.25
S4. Using Assessment Results to Make Decisions			
Q4 …inappropriate uses of assessment	-.02	-	.23
Q11 …deciding on student instructional needs	.22	.21	-.02
Q18 …administering pre-tests for instruction	.25	.20	.07
Q25 …purpose of formative assessment	.19	.20	.19
Q32 …discrepancies with standardized tests	.16	.19	.13
S5. Developing Valid Grading Procedures			
Q5 …steps to improve grading procedures	.22	.17	-
Q12 …grades least reflective of achievement	.25	.19	.20
Q19 …criticism of grading based on tests only	.19	.19	.21
Q26 …consistency in grading practices	.16	.19	.29
Q33 …grading systems based on content mastery	-.08	.20	.18
S6. Communicating Assessment Results			
Q6 …best explanation of student grades	.25	.21	.28
Q13 … comparing percentile ranks	.19	-	-
Q20 …distinguishing between grading systems	.16	.19	.16
Q27 …using grade equivalents	-.08	.18	.26
Q34 …explaining percentile ranks	-.08	.20	.23
S7. Recognizing Unethical or Illegal Practices			
Q7 …appropriate use of assessment information	.19	.19	.23
Q14 …strategies during statewide assessments	.16	.19	.13
Q21 …adjusting test scores to improve grades	-.08	.21	.14
Q28 …best practices for clarifying test items	-.08	.20	.21
Q35…violationsofstudentinformationpolicies	.01	.18	.28

Final Solution for BEL, PRA and ALI

Overall, factor and item analyses yielded three reduced item measures for assessment beliefs (BEL), assessment practices (PRA), and assessment knowledge (ALI). As shown in Table 4.26, the BEL measure retained 11 of the 14 original items for pooled mean of 32.91 (*SD* = 5.83) with good reliability among the remaining BEL items (α = .86). The PRA measure retained 17 of the 18 original items for pooled mean of 32.91 (*SD* = 5.83) with excellent reliability among the remaining PRA items (α = .90). The ALI measure retained 27 of the 35 original items for pooled mean of 15.85 (*SD* = 3.95) with questionable reliability among the remaining ALI items (α = .63). Averaged skewness and kurtosis across imputed datasets are within range across all measures, assuming a normal distribution of items. The retained items, along with the assessment learning experiences variables were used to examine the assessment leadership model.

Table 4.26

Pooled Mean Scores for ASLS Measures

Variable	# Items	*M (SD)*	Skewness	Kurtosis
Beliefs (BEL)	11	32.91 (5.83)	-.40	.35
Practices (PRA)	17	54.91 (22.25)	-.32	-.15
Knowledge (ALI)	27	15.85 (3.95)	.09	-.50

Research Question Findings

The first research question (RQ1) was: *To what degree do assessment learning experiences, beliefs, and knowledge influence the assessment leadership practices of school and classroom leaders?* To examine this model, multiple regression (MR) with maximum likelihood (ML) estimator was conducted using the 100 imputed datasets. First, the correlation matrix was analyzed to determine the relationships among variables. As shown in Table 4.27, variables had low correlations, ranging from *r* = .01 (assessment

knowledge and learning experiences) to r = .34 (postsecondary courses and professional development sessions), which suggested multi-collinearity was not a problem. Correlation coefficients between independent and dependent variables also were small.

Table 4.27

Correlation Matrix for Original Model

Variable	PRA	EPS	EPD	BEL
EPS	.06			
EPD	.13	.34		
BEL	.32	.20	.32	
ALI	.20	.01	.01	.05

Note. PRA=assessment practices; EPS=assessment learning experiences in postsecondary courses; EPD=assessment learning experiences in professional development sessions; BEL=assessment beliefs; ALI=assessment knowledge.

Results from the multiple regression analysis indicated that the 4-predictor model was statistically significant, R^2 = .15, $F(4, 279) = 11.92$, $p < .001$. Both assessment beliefs (β = .06, z = 5.24, $p < .001$) and assessment knowledge (β = .07, $z = 2.68$, $p = .01$) evidenced a positive slope with assessment practices. Table 4.28 shows the estimates variables in the 4-predictor model with associated *z*-scores and *p*-values.

Table 4.28

Multiple Regression Results with 4-Predictor Model on Assessment Practices

Predictor	*b*	β	*z*	*p*
EPS	.08	.07	1.06	.29
EPD	.00	.07	-0.06	.95
BEL	.32	.06	5.24	.00
ALI	.18	.07	2.68	.01

Note. EPS=assessment learning experiences in postsecondary courses; EPD=assessment learning experiences in professional development sessions; BEL=assessment beliefs; ALI=assessment knowledge.

The second (RQ2) and third (RQ3) research questions were: *To what extent does school role (i.e., principal, assistant principal, or PLC facilitator) moderate the relationship between assessment learning experiences, beliefs, and knowledge and*

assessment leadership practices of school and classroom leaders? To what extent does school type (i.e., elementary or secondary) moderate the relationship between assessment learning experiences, beliefs, and knowledge and assessment leadership practices of school and classroom leaders? First, a second multiple regression analysis was conducted with the additional variables. As mentioned, the original 4-predictor model had an R^2 = .15, $F(4, 279) = 11.92$, $p < .001$, with assessment beliefs and assessment knowledge having significant positive slope with assessment practices. The expanded 6- predictor model including school role and school type as predictors had an $R^2 = .17$, $F(6, 277) = 9.26$, $p < .001$. Both assessment beliefs ($\beta = .06$, $z = 5.05$, $p < .001$) and assessment knowledge ($\beta = .07$, $z = 2.69$, $p = .01$) retained significant positive slope on assessment practices with school type having significant negative slope on assessment practices ($\beta = .06$, $z = 5.24$, $p < .001$). The additional preditor in the 6-predictor model suggests that elementary school and classroom leaders tended to report engaging in more assessment practices than secondary school and classroom leaders. Table 4.29 shows the estimates for each of the variables in the 6-predictor model with associated *z*-scores and *p*-values.

Table 4.29

Multiple Regression Results with 6-Predictor Model on Assessment Practices

Predictor	*b*	β	*z*	*p*
TYP	-.14	.06	-2.40	.02
ROL	.06	.06	1.03	.30
EPS	.09	.07	1.25	.21
EPD	.01	.07	0.09	.93
BEL	.31	.06	5.05	.00
ALI	.18	.07	2.69	.01

Note. TYP=school type; ROL=school role; EPS=assessment learning experiences in postsecondary courses; EPD=assessment learning experiences in professional development sessions; BEL=assessment beliefs; ALI=assessment knowledge.

Next, a hierarchal multiple regression with maximum likelihood estimator was calculated using the 100 imputed datasets to determine whether the assessment leadership model varied due to school role or school type. As shown in Table 4.30, low to moderate correlations were observed for most variables, ranging from $r = .00$ (school role by professional development and school type by knowledge) to $r = .66$ (school role by postsecondary courses and school role). Like the original model, there were minimal correlations among variables, suggesting multi-collinearity was not a problem. Five correlations were above $r > .70$, likely due to containing common interaction variables.

Table 4.30

Correlation Matrix for Full Model

Variables	PRA	TYP	ROL	TYP*EPS	TYP*EPD	TYP*BEL	TYP*ALI
TYP	.13						
ROL	.08	.09					
TYP*EPS	.04	.39	.02				
TYP*EPD	.00	.38	.05	.51			
TYP*BEL	.08	.97	.10	.38	.36		
TYP*ALI	.06	.94	.12	.36	.36	.92	
ROL*EPS	.06	.12	.66	.13	.02	.09	.16
ROL*EPD	.40	.01	.60	.02	.01	.00	.00
ROL*BEL	.04	.08	.98	.04	.05	.11	.11
ROL*ALI	.01	.09	.96	.01	.05	.10	.15
EPS	.06	.09	.08	.76	.34	.09	.08
EPD	.13	.10	.03	.34	.80	.09	.09
BEL	.32	.06	.06	.02	.04	.11	.04
ALI	.20	.07	.14	.05	.02	.08	.31

Note. PRA=assessment practices; TYP=school type; ROL=school role; EPS=assessment learning experiences in postsecondary courses; EPD=assessment learning experiences in professional development sessions; BEL=assessment beliefs; ALI=assessment knowledge.

The full model including interactions of school role and type had an $R^2 = .18$, $F(14, 269) = 4.28$, $p < .001$, with only assessment beliefs still having significant positive slope with assessment practices ($\beta = .09$, $z = 4.04$, $p < .001$). When compared to the

nested model, the results suggested that the interactions of school role and school type did not statistically improve the 6-factor model, $\Delta R^2 = .02$, $F(8, 269) = .66$, $p = .73$.

Ancillary Analyses

An additional variable was introduced to explore whether the expansion of the assessment learning experiences definition to include the frequency of PLC meetings survey item influenced the model. As discussed, PLCs often serve as job-embedded professional development and could be argued contribute to assessment learning experiences, especially given the nature of PLC activities in the target district. First, multiple regression (MR) with maximum likelihood (ML) estimator was conducted using the 100 imputed datasets with the original variables and the addition of the PLC variable, yielding a 5-predictor model. The results indicated that the 5-predictor model was statistically significant, $R^2 = .17$, $F(5, 278) = 11.07$ $p < .001$. Like assessment beliefs and knowledge, frequency of PLC meetings ($\beta = .06$, $z = 2.51$, $p = .01$) also evidenced a positive slope with assessment practices. The findings suggest that the more often school and classroom leaders reported attending PLC meetings, the more often respondents reported engaging in assessment leadership practices. Table 4.31 shows the estimates for each of the variables in the 5-predictor model with associated *z*-scores and *p*-values.

Table 4.31

Multiple Regression Results with 6-Predictor Model on Assessment Practices

Predictor	*b*	β	*z*	*p*
PLC	.14	.06	2.51	.01
EPS	.07	.07	0.94	.35
EPD	.00	.07	0.07	.95
BEL	.32	.06	5.29	.00
ALI	.17	.07	2.49	.01

Note. PLC=frequency of PLC Meetings; EPS=assessment learning experiences in post-secondary courses; EPD=assessment learning experiences in professional development sessions; BEL=assessment beliefs; ALI=assessment knowledge.

Next, a hierarchical regression analysis was conducted with PLC as an explanatory variable. The full model including interactions of school role and type had an $R^2 = .21$, $F(17, 266) = 4.26$, $p < .001$, with assessment beliefs ($\beta = .09$, $z = 4.04$, $p < .001$) as the only variable retaining significant positive slope with assessment practices. Similar to the 6-predictor model, when compared to the nested model, the results suggested that the interactions of school role and school type did not statistically improve the 5-factor model, $\Delta R^2 = .04$, $F(12, 266) = 1.24$, $p = .26$. In this study, the assessment leadership model did not vary significantly based on school role or school type.

Summary

Chapter 4 outlined the results of the study based on the research questions and design. Response rates for the survey were lower than anticipated and included significant missing data. Thus, a series of missing data analyses were conducted to resolve missing data issues. Factor and item analyses suggested marginal to poor fit, even for measures that had shown better fit in previous studies. Several steps were taken to identify the best factor structure given loadings and reliability statistics. The results of the multiple regression analyses suggested that assessment beliefs and knowledge were significant predictors of assessment practices, as well as school type when the model was expanded to include six predictors. Assessment learning experiences as defined in this study did not yield significant results. Ancillary analyses suggest frequency of PLC meetings, specifically if centered on job-embedded professional development in assessments, may be considered as part of an expanded definition of assessment learning experiences. When compared to the nested models, the model did not vary by school role

or type when interactions were incorporated into the analyses. However, these results should be interpreted with caution given limitations due to new item construction, self-report, administration on multiple devices (i.e., computer, mobile) smaller sample size in comparison to the target population, and significant missing data. The results of this study underscore the importance of continued research in assessment leadership, specifically if survey methods are utilized to reduce bias and increase power. Chapter 5 summarizes the findings and makes recommendations for future studies.

CHAPTER 5

DISCUSSION

Over the last 50 years, schools have been expected to meet the increasing demands of accountability policies. In response to heightened expectations for student achievement outcomes, educational practices also have advanced to incorporate innovative approaches to curriculum, instruction, and assessment (Darling-Hammond, 2004; Fullan, 2001; Green, 2010; Noonan & Renihan, 2006). Unfortunately, despite the expanded emphasis on effective educational practices in policy and practice, educators have faced many challenges to designing and implementing school-renewal efforts such as sufficient time, access to resources, and effective professional development (Stiggins & Duke, 2008; Volante & Cherubini, 2011). As a result, principals have explored collaborative instructional leadership approaches that activate teachers as leaders to inform and support implementation, through transformational, distributive, and collective leadership models (Collins, 2001; Green, 2010; Hallinger & Heck, 2010; Lewis et al., 2010; Leithwood, 1992; Noonan & Renihan, 2008; Stiggins & Duke, 2008).

Assessment Leadership Defined

Central to reform efforts are evolutions in student assessment systems (Gallagher et al., 2008; Kingston & Nash, 2011; Stiggins & Duke, 2008). Traditional models of student assessments rely on grades from summative assessments as the primary measures of student achievement (Ingram et al., 2004). More contemporary models incorporate multiple forms of student assessments collected prior to, during, and after instruction. Data are used not only to determine final grades but also to provide teachers with information about student progress toward achieving academic standards throughout the

school year. Conducting ongoing assessments of student data assists teachers in making strategic adjustments to curriculum and instruction with the goal of matching the content and delivery to student needs. Effectively, evolutions in student assessment systems transform teachers from merely being deliverers of content to serving as responsive facilitators of student learning. While this approach has been purported to translate into positive student-learning outcomes (Coggshall et al., 2012; Hattie 2009; Robinson et al., 2008; Valentine & Prater, 2011), schools struggle with implementation on many fronts.

Foremost, schools have limited time or access to a variety of student assessment tools. In addition, principals and teachers report lacking the assessment knowledge and skills necessary to develop and execute robust student assessment systems. In response, many schools have implemented and embraced professional learning communities (PLC) as forums for discussing student data and planning instruction based on evidence gathered. PLCs provide principals with an avenue through which to distribute leadership among teacher leaders as well as position teachers as leaders among their peers (Danielson, 2009; Dufour & Eaker, 1998). PLCs assume school leaders have both established a vision and created conditions for shared decision making among stakeholders, including providing them with the requisite time and tools.

Assessment leaders are critical to successful implementation of student assessment systems within PLCs or other collaborative leadership frameworks, and ultimately, contribute to improvement efforts aimed at enhancing student achievement. Assessment leadership is defined as instructional leaders who (a) establish a vision that sets clear and appropriate expectations for student assessment systems, (b) lead data discussions, (c) foster assessment literacy in teachers through ongoing collaborative

learning experiences, and (d) self-reflect on personal assessment practices (Brookhart, 2001; Noonan & Renihan, 2008). Assessment practices are influenced by three key components of assessment leadership: assessment learning experiences, beliefs, and knowledge. These components, along with the setting events that lay the foundations for assessment work, comprise the conceptual framework for study in assessment leadership.

Methodology Summarized

A nonexperimental correlational research design was utilized to answer research questions concerning assesment leadership preparation and practices among school and classroom leaders in a selected large school district in Florida, which had dedicated several years to providing practice-based professional development related to assessment leadership. Data were collected using an Internet-based survey constructed by the researcher from a principal reflection tool aimed at capturing key components of assessment leadership (Noonan & Renihan, 2008), coupled with an existing measure of assessment literacy, *Assessment Literacy Inventory* (ALI) (Mertler & Campbell, 2005). Additional questions pertaining to leader background and assessment learning experiences were also included in the survey. A series of data analyses were conducted, starting with missing data analysis. Significant patterns of missing data were identified and handled through multiple imputation methods prior to statistical analyses.

After imputation, descriptive analyses of background, demographics, and pooled item means and standard deviations were calculated. Next, factor analysis paired with classical item analysis were used to examine the factor structure of the assessment beliefs, practices, and knowledge measures. Last, multiple regression analyses were conducted with the revised measures to examine the predictive qualities of the assessment components on practices. Hierarchal regression analysis also was completed to

determine if the models varied based on school role or type. In addition to the quantitative analyses, documents from the target district were reviewed to describe the contextual conditions for assessment leadership practices in the target school district. The findings are framed within the context of educational reform and professional development efforts in the target school district.

Assessment Leadership Model Outcomes and Implications

Over the past four years, district and school administrators implemented a comprehensive PLC model whereupon teacher leaders were trained to facilitate instructional discussions with teachers on a weekly basis. This structure provided teachers the dedicated time and resources to review student data and plan instruction based on student progress toward standards achievement. In addition, teacher teams utilized this time to develop common student assessments as a standardized measure of student-learning outcomes. These tools provided school and classroom leaders with a common assessment system aligned to a data decision protocol, which were critical to successful PLC implementation (Danielson, 2009; Dufour & Eaker, 1998).

It is clear from a comprehensive review of the documents that the school district planned for the foundational tenets of assessment leadership. District personnel created a vision statement rooted in data-informed decision making to inform instruction and underscored by collaborative cultures. District and school leaders actualized the vision with clear expectations for leveraging PLC members to engage in assessment practices. The infrastructures were observable and linked to the district vision. In addition, school and classroom leaders were provided with extensive and ongoing professional development supported with instructional and assessment resources to prepare them with the knowledge and skills necessary to implement the model successfully. The district

clearly was committed to ensuring school personnel understand the expectations and are provided with tools to support implementation.

Assessment Beliefs on Assessment Practices

Multiple regression analyses of survey data revealed two major findings. The most robust predictor of reported assessment practices was assessment beliefs, defined as *attitudes that assessment practices are essential components to school-improvement efforts*. In this study, respondents who reported the strongest assessment beliefs also tended to have the highest reported engagement in assessment practices. Across the beliefs measure items, school and classroom leaders shared the most agreement concerning using classroom and student context to assess student learning and to inform decisions about what students know and can do. Moreover, they indicated that student assessment data are primary sources for understanding what students know and can do as well as what to teach and how to teach. These outcomes support the notion that assessment beliefs are positively related to assessment practices and reinforce the need for district leaders to understand the belief systems of school and classroom teachers when implementing new initiatives.

Biases about assessments and data use can create obstacles to effective implementation (Burke & Wang, 2010; Coburn & Talbert, 2006; Deenen & Brown, 2016; Gallagher et al, 2008; Kerr et al., 2006; Popham, 2010; Webb, 2002). In fact, when school and classroom teachers question student assessment outcomes, they tend to resort to traditional practices of delivering content and informally monitoring student learning (Young & Kim, 2010). School and classroom leaders improve assessment beliefs by creating a vision and establishing shared decision making within the framework of PLCs that build trust among members (Danielson, 2009; Dufour & Eaker,

1998). Moreover, assessment leaders influence assessment beliefs by gaining consensus, creating reliable student data systems, and talking about student assessment data in relevant ways (Boudett et al., 2010; Dufresne & McKenzie, 2009; Leithwood & Louis, 2012). District leaders should take steps to prevent issues related to assessment beliefs by evaluating the current assessment beliefs of their school and classroom leaders and then strategically planning effective strategies to address needs.

School and classroom leaders were clear that student and classroom contexts are as important to learning and teaching as putting in place comprehensive assessment systems that monitor student progress toward standards. Moreover, both school and classroom leaders reported disagreeing with the notion that teachers engaging in assessment practices and leaders promoting effective assessment practices are as important as teaching. Even though these outcomes were not surprising, given that the respondents were teachers and negative perceptions around standardized assessments have evolved in the last decades, these assessment belief items require further exploration. Further, the outcomes suggest school and classroom teachers have a greater appreciation for the qualitative aspects of student-learning situations than standards-based student assessments (Noonan & Renihan, 2008) as well as for the art of teaching versus the art of assessing (Guskey, 2009; Marzano, 2006; Welch et al., 2007). Schools should consider expanding current assessment models to systematically incorporate both qualitative and quantitative information ways that are consumable by PLCs and other forums where instructional decisions are made. Moreover, district leaders should work to understand school and classroom leaders' perceptions of assessments as they relate to teaching and target professional development to the explicitly reinforce the connection with data protocols to inform implementation (Boudett et al., 2010).

The target school district for this study allocated considerable time and resources over the past years in developing consensus around their assessment-related initiatives. District personnel created administrator and teacher resources that clearly stipulate expectations for data-informed decision making and collaborative school cultures. Moreover, they provided ongoing professional development aimed at establishing a culture of inquiry and monitored outcomes through walkthroughs and feedback systems. The significant finding between assessment beliefs and practices is not surprising, given the time and attention district personnel dedicated to cultivating conditions for developing assessment beliefs. However, given the outcomes and the relative emphases on classroom and student context as measures of learning, the school system may consider putting in place systems for assessing and monitoring beliefs as they evolve their initiatives to ensure their leaders remain on track with the vision.

Assessment Knowledge on Assessment Practices

The other major finding was that assessment knowledge significantly predicted reported assessment practices. Assessment knowledge was defined as *competencies necessary to be assessment literate.* In this study, respondents who obtained higher total scores on the *Assessment Literacy Inventory* tended to report engaging in more frequent assessment practices. The 27-items retained in the final inventory measure spanned all seven assessment literacy standards; however, most items fell into four main categories: (a) recognizing ethical practices, (b) developing valid grading practices, (c) using assessments to make decisions, and (c) communicating assessment results. School and classroom leaders' performance on these standards is consistent with previous studies using the inventory for ethical and grading practices and creating and using student

assessments to inform instruction (Davidheiser, 2013; Hameister, 2013; Matthews, 2007; Perry, 2013). The relationship between demonstrated assessment knowledge on the ALI and reported assessment practices suggest the continued importance of school districts developing administrators and teachers as assessment literate leaders.

Acting ethically is one of the six core components of the Interstate School Leaders Licensure Consortium standards for educational leaders (Saunders & Kearney, 2008). Ethical behaviors entail developing a vision and aligning instructional practices with ethical guidelines, not unlike components of assessment leadership (Dufresne & McKenzie, 2009). Awareness of and knowledge about ethical policies assists leaders in adhering to ethical behaviors. The *Assessment Literacy Inventory* outcomes for principals and teachers in this study are consistent with previous studies. On average, leaders tended to perform better on ethical assessment practice items than on items dealing directly with student data such as administering and interpreting assessments (Davidheiser, 2013; Impara & Plake, 1995; Perry, 2013). District leaders should consider these outcomes when planning professional development activities for their employees. School and classroom leaders would benefit from more emphasis on the application of assessment-related ethical behaviors than on theories and policies of ethical practices. This was evident among school and classroom leaders in this study reporting engaging in assessment activities such as creating assessments and organizing data into graphs and charts less often than other assessment practices.

Developing valid grading practices is a one of the main tenets of the *Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA, 1990). Competencies in this area are essential to learning and teaching practices.

Movements centered on standards-based grading practices underscore the importance of this assessment literacy competency, namely that teachers use valid and reliable assessment strategies to communicate grades with students and parents (Brookhart, 2011, Muñoz & Guskey, 2015). Even though leaders in the district evidenced relatively strong assessment knowledge in this area, defining what valid and reliable grading practices are in classrooms continues to be a problem of practice. Unfortunately, many teachers still rely on traditional approaches to assess students and assign grades (Penuel et al., 2004; Roschelle et al., 2005). Continued emphasis on grading systems and competencies are at the forefront of taking steps toward more valid and reliable grading procedures.

Using assessments to make decisions and communicate outcomes in consumable and relevant ways are central competencies of assessment leadership. However, competencies in these areas are not independent of selecting appropriate assessment tools and organizing and interpreting assessment data, areas in which school and classroom leaders tended to exhibit fewer proficiencies. In fact, the first step to effective data-informed decision making is selecting the most appropriate tool to answer the question or problem at hand (Burke & Wang, 2010; Coburn & Talbert, 2006; Gallagher et al, 2008; Kerr et al., 2006) Moreover, using multiple assessment methods together enhances leaders' abilities to understand the reasons underlying the data to make the most effective instructional decisions (Boudett et al., 2010; Black et al., 2004; Lee & Wiliam, 2005; Stiggins, 2002; Suurtamm et al., 2010; Young & Kim, 2010). Unfortunately, teachers report valuing formative assessments over summative assessments, but they are more confident with summative assessments to inform decisions (Henry, 2011; Penuel et al., 2004; Roschelle et al., 2005) and tend to rely on traditional approaches (Young & Kim,

2010). Given the outcomes of this study, coupled with research emphases in these key assessment literacy competency areas, teacher educators and district leaders should develop proficiencies across all seven assessment literacy standards. Specifically, training should focus on creating, administering, and interpreting assessments that broadens their effectiveness to make decisions and communicate results are impacted.

The target school district in this study has made strides in building assessment knowledge; however, upon review of their professional development activities, assessment literacy has been secondary to other competencies such as instructional delivery models and content-area resources and standards. One might argue that skills in data-informed decision making provide the foundation for making decisions about delivery and resources, especially given the finding that school and classroom leaders value classroom and student context when making decisions and assessing student-learning outcomes. Although the target district has implemented a myriad of systems and resources, even with their efforts to date, the research is clear that school and classroom leaders still experience challenges to beliefs and competencies related to assessment leadership (Gallagher, Means, & Padilla, 2008; Means et al., 2009; Volante & Cherubini, 2011; Young & Kim, 2010). Thus, continued professional development opportunities in response to school and classroom leaders' needs are necessary to ensure that (a) there is not drift from the central vision and (b) teachers continue to improve assessment knowledge across all seven competency standards.

Assessment Learning Experiences on Assessment Practices

Unlike assessment beliefs and knowledge, assessment learning experiences did not significantly predict reported assessment practices as defined by the

assessment leadership framework created for this study. Assessment learning experiences were defined as *the number of assessment-related learning opportunities experienced during preservice and inservice training*. In this study, respondents were asked to report the number of postsecondary courses and inservice professional development sessions they completed. Despite the insignificant findings on assessment practices, the research is clear about the importance of assessment learning experiences for establishing effective assessment leadership practices. In fact, principals and teachers reported feeling underprepared and lacking the pre-requisite training in assessment leadership practices (Clifford & Mason, 2013; Ulmer, 2002).

Administrators and teachers experience few university-level courses focused on assessment leadership, which could not only impact their assessment knowledge but also influence their assessment beliefs and experiences with assessments (Popham, 2010; Stiggins, 2001; Wayman, Midgley & Stringfield, 2006). Respondents in this study indicated experiencing more than four times the number of inservice professional development sessions than preservice postsecondary coursework. Moreover, they reported up to 150 sessions, which reflects an extensive amount of training. It is possible that the comprehensive training approached offered in the target district over the last four years neutralized the relative influence of this variable on any outcomes. Future research in assessment learning experiences should define and measure more thoroughly this variable as a component of assessment leadership models, especially given the fact that the few individuals that responded to the final question indicated not feeling prepared for the role.

It should be noted that although the number of PLC meetings was not part of the assessment learning experiences variable nor factored into the original assessment

leadership model, ancillary analysis suggested that number of PLC meetings was a significant predictor of assessment practices. The more often school and classroom leaders attended PLCs (i.e., weekly, bi-weekly, monthly, quarterly), the higher their reported engagement in assessment leadership practices. This finding is not surprising, given the clear vision, structures, and professional development provided by the target district to support weekly PLCs. This result may suggest circular logic in that the more school and classroom leaders attend an event, the more often they would report attending the event. However, in this case, it also could be argued that (a) the outcomes validate the target districts efforts to implement the PLC model and (b) the PLCs offer a viable forum for engaging in assessment practices. Both conclusions, although not by design, underscore the importance of continuing to implement assessment practices within the PLC model. Further, the definition of assessment learning experiences may need to expand to include other factions of experience beyond inservice and preservice education.

School Role and Type

In this study, variables of school role or type did not significantly vary the assessment leadership model. Administrators and teachers at the elementary and secondary levels shared similar outcomes for assessment learning experiences, beliefs, and knowledge on reported assessment practices. However, there were significant differences observed when comparing means. Of note, school and classroom leaders at the elementary level reported engaging in more assessment practices for setting a vision, conducting data discussions, and organizing data into graphs and charts for use to inform decisions compared to school and classroom leaders at the secondary level. This also was evident in the 6-factor model, which suggested that school type was a significant

predictor of assessment practices. These outcomes may be attributed to the elementary school structures where teachers tend to share physical classroom workspaces and more readily loop students within and across grade levels, which are unlike secondary school structures where teachers typically are organized by content areas. The fluidity of elementary school environments may more easily lend itself to authentic collaborative opportunities. To further explore the assessment leadership model by school role and type, future studies should expand the variables to individual grade levels and other role types (i.e., reading specialist, assistant principal, principal). By adding more dimension, differences may better explain how the model functions by school role and school type.

Limitations

First, despite the large size of the target population and number of opportunities to complete the survey over the 6-week period, the response rate was smaller than anticipated. The lower response rate resulted in a small sample for the planned multi-level modeling. While the district was in support of the administration time period, the survey was administered toward the end of the school year and participants were solicited via electronic mail on a volunteer basis only. These factors may have limited the number of respondents for a number of reasons, such as seasonality, survey length, administration mode (i.e., computer or mobile), interest, or motivation to complete the survey. Moreover, the respondents that completed the survey may have been more motivated than others due to their support or nonsupport for assessment-related initiatives in the district. This factor could have skewed the outcomes in positive or negative directions. The use of computer compared to mobile device may have made it easier to take the survey, which also could have impacted motivation and persistence to complete the

survey. Moreover, this study was conducted in one school district in Florida, and the target population did not include a control group or randomization in participant selection, which presents additional threats to external and internal validity, respectively.

In addition to the small sample size, there were significant missing data in this study. Even though field-test participants completed the survey within the expected 30 to 60-minute timeframe, during the live survey, several participants stopped and did not return to complete, resulting in partial completions. The monotone-like response pattern resulted in progressively more missing data for the last four sections of the survey, of which three focused on assessment leadership. Respondents may have failed to persist to the end of the survey due to the length of the survey or they stopped, intending to return, but failed to complete due to competing activities, interests, or ease of administration, depending on the type of device that was selected to complete the survey. Possible reasons for general patterns of missing data may have been difficulties with item content, complexity or technology (i.e., the responses did not save properly). A few of the respondents noted in the open-ended question that the knowledge section was difficult and "felt like a test," which suggests perceived difficulty or impact on motivation. Multiple imputations were conducted to correct for the missing data. Although multiple imputation is a viable approach to handling missing data, with the level of missing data in this study, outcomes should be interpreted with caution and validated with replication.

Finally, the survey consisted of primarily newly constructed items and thus posed threats to both reliability and validity. Even though this study was designed as exploratory, given that no comprehensive surveys of assessment leadership have been established, the small sample size and missing data only compounded what were

anticipated limitations in new instrument development, which made it difficult to evaluate the structure. Several methods were employed to analyze the stability, dimensionality, and reliability of the factors (i.e., confirmatory factor analyses, exploratory factor analyses, classical item analysis) and used together to inform the final structures. However, it should be noted that the fit was marginal to poor, suggesting a weak factor structure, and reliability analyses were not as robust as preferred when validating scales. The analysis procedures and outcomes in this study set the stage for future research in both assessment leadership and methods for data collection and measurement models. Finally, the assessment practices were reported and not observed, which may also have skewed the actual representation of engagement.

Implications for Future Research and Practice

Since this study was intended to be exploratory in nature, the outcomes of the *Assessment Leadership Survey* provide a foundation for future research in assessment leadership. It is recommended to revise the survey based on outcomes of the factor and item analyses; thus, revisions should center on item construction and organization. First, items with low loadings should be reviewed and revised to improve clarity. In addition, assessment learning experiences as a construct should be reexamined to consider incorporating other components beyond number of trainings. The item response formats for these items also should be revisited to consider changing from entering numbers to selecting ranges. Second, it is recommended to ungroup the items that currently are organized into testlets. Even though unbundling items creates redundancy among item stems, this approach would increase opportunities for independence among items and randomization within the measures. Last, the scenarios on the *Assessment Literacy*

Inventory should be updated to match more contemporary examples since the most recent revision was completed over a decade ago (Mertler & Campbell, 2005). Although it was not in the scope of this study to revise the inventory, many of the item steps and responses contain excessive words. Moreover, there have been evolutions in educational practices that could inform improved scenarios.

Further, it is recommended to alter the way the survey is administered as well as to consider parallel forms to evaluate the best method. The current survey was organized into five sections within randomization within each block. Two alternative administrations could be considered: (1) block randomization or (2) elimination of sections. Block randomization would provide a small adjustment to the current administration whereupon the items would still be organized into sections and randomized within blocks. However, the presentation of the blocks themselves would be randomized. The other option is to detach items from sections and create complete randomization of items throughout the survey; however, there are advantages and disadvantages to this approach. The advantage is that the survey items would be completely randomized and thus reduce order bias. Given the differences in item construction and response types (i.e., multiple-choice, agreement Likert-scale, frequency Likert-scale, correct-incorrect) across assessment leadership variables, respondents may get confused or frustrated by the varying directions or frustrated or styles. For these reasons, studies designed to analyze outcomes using parallel forms would help inform the best approach to survey administration.

Next, future survey administrations should consider alternative delivery methods. One consideration is planned missing data patterns (Enders, 2010). Given the length of

the survey and the importance of each component to the overall model, the survey could be administered to the target population in reduced portions. For this approach to be effective, the research design must include a large sample and planned multiple imputation methods as part of the data analysis procedures. Since missing data are an unfortunate but known reality of survey research (Enders, 2003; Peng et al., 2006), planned missing data patterns may provide an avenue to collect extensive amounts of data, and increase response rates, while also attending to known issues in survey research. In addition, studies may consider tracking the type of device used (i.e., computer or mobile) to determine if it influenced the survey completion or outcomes.

Once the factor structure of the tool has been validated, the assessment leadership practices of various populations should be examined. The target population for this study was intentionally selected because of the district's recent initiatives in assessment-related activities within a PLC framework. Thus, respondents may have been more likely to demonstrate assessment leadership practices compared to other districts that have not been engaged in this level of systematic reform. To add more dimension to the research design, future research in assessment leadership should expand the target population to include teachers not in designated leadership roles but who are influenced by the assessment leadership practices of others. One of the goals of a PLC is to transfer knowledge among teachers (Danielson, 2009; Dufour & Eaker, 1998) with the hope that teachers assume the role of assessment leaders of their classroom (Guskey, 2009). Moreover, the *Assessment Leadership Survey* could be used to compare districts with differing assessment-related implementation models and states with differing accountability policies and expectations for practice. The research design could

investigate the model using a nested structure approach, which evaluates the survey responses relative to school leaders, teacher leaders, and teachers engaged in assessment leadership practices within the same school buildings as compared to other buildings.

Assessment leadership practices have been purported to lead to positive student outcomes (Boudett et al., 2010; Leithwood & Louis, 2012). However, no known studies have examined the impact of these behaviors on student achievement. Future studies that investigate assessment leadership components as it relates to both assessment practices and student outcomes are essential to understanding what practices are most effective for students. Research designs that incorporate both quantitative and qualitative methods should be explored to balance the survey results with conversation about leader experiences. The research design could incorporate observations of assessment leadership behaviors; interviews with school and classroom leaders, teachers, and students; and focus groups. A mixed-methods approach would provide context to the results and add to the interpretation (Creswell, 2009). The brief insights gleaned from the final question with respect to the assessment leadership questions provided the basis for next steps. Future studies should continue to refine the assessment leadership framework to incorporate additional components as informed by practitioners engaged in this work.

Last, students self-monitoring their own progress toward academic standards has received more attention (Andrade & Valtcheva, 2009; Black et al., 2004; Hattie, 2009; Welch, Adams et al., 2007; Stiggins, 2005). In fact, students tracking their performance on assessments has evidenced an additive impact on student achievement to teachers tracking alone, especially when graphical displays are used (Marzano, 2010). Future models of assessment leadership should explore the relative influence of students as

assessment leaders in the classrooms. Such investigations also should include examinations among assessment beliefs, knowledge, and practices of teachers and students and the relative impact of these behaviors on student-learning outcomes.

Conclusion

Educational policies aimed at student achievement have been and will continue to drive shifts in educational research and practice. At the center of shifts are educational leaders, individuals responsible for interpreting policies and planning adjustments to practices based on what has been proven effective in research. Thus, school reform hinges on the abilities of school and classroom leaders to understand the problems of practice at hand and design strategies in response to identified needs. Skills in data-informed decision making are essential to leaders successfully navigating school improvement processes. Unfortunately, leaders face challenges developing skills in this area, including time, access to resources, beliefs about data, and requisite knowledge for collecting, organizing, and using data to inform decisions. Although schools have made strides, such as with vision, time and resources, comprehensive student assessment systems and embedded professional learning opportunities through PLCs, to address these barriers, more work is needed. A continued emphasis in research on assessment leadership practices can help inform the conditions necessary to accomplish school assessment reform centered on using data to make educational decisions.

Appendix A

Electronic Consent

Dear school or classroom leader of Pasco County Schools,

You are invited to participate in a web-based survey on assessment leadership. This survey is a research project being conducted by Carrie Morris, a doctoral candidate in the Department of Educational Leadership Studies at the University of Kentucky. Your responses may help Pasco County Schools better understand leaders' assessment learning experiences, beliefs, and knowledge to inform professional development and other leadership supports. Please note that while the survey is adapted for hand-held devices, you may find it easier to take it on a computer. I ask that you respond to the best of your abilities based on your current knowledge and experiences in this area at this time.

We hope to receive completed surveys from a minimum of 600 people, so your answers are important to us. Of course, you have a choice about whether or not to complete the survey, but if you do participate, you are free to skip any questions or discontinue at any time. The survey should take approximately 30 minutes to complete. There are no known risks to participating in this study. Your response to the survey is anonymous, which means no names will appear or be used on research documents, or be used in presentations or publications. Your survey answers will be sent to a link at Qualtrics.com where data will be stored in a password protected electronic format. No one will be able to identify you or your answers, and no one will know whether or not you participated in the study.

If you have questions at any time about the study or the procedures, you may contact my research supervisor, Professor Dr. Tricia Browne-Ferrigno by phone at (859) 257-5504. If you feel you have not been treated according to the descriptions in this form, or that your rights as a participant have not been honored during the course of this research project, or you have any questions, concerns, or complaints that you wish to address, you may contact the Office of Research Integrity by mail at 315 Kinkead Hall**,** University of Kentucky**,** Lexington, KY 40506-0057 or by phone at (859) 257-9428.

To provide your electronic consent, please select your choice below. You may print a copy of this consent form for your records. Clicking on the “Agree” button indicates that

- You have read the above information
- You voluntarily agree to participate
- You are 18 years of age or older

☐ Agree
☐ Disagree

Appendix B

Assessment Leadership Survey (ASLS)

Section 1: Education Background

For the first set of questions, provide your best responses.

Q1 What grades do you currently teach or lead as an administrator? Check all that apply.
- ☐ PK
- ☐ K
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7
- ☐ 8
- ☐ 9
- ☐ 10
- ☐ 11
- ☐ 12

Q2 What is your current school LEADERSHIP ROLE?
- ❍ Principal
- ❍ Assistant Principal
- ❍ PLC facilitator

Q3 How many years' experience do you have as a CLASSROOM TEACHER?
- ❍ 1-5 years
- ❍ 6-10 years
- ❍ 11-15 years
- ❍ 16 or more years

Q4 How many years' experience do you have as a PROFESSIONAL LEARNING COMMUNITIES (PLC) FACILITATOR?
- ❍ none
- ❍ 1 year
- ❍ 2 years
- ❍ 3 years
- ❍ 4 years

Q5 How many years' experience do you have as a SCHOOL ADMINISTRATOR? (Enter an integer) ___

Q6 How many POSTSECONDARY COURSES have you taken on CLASSROOM AND/OR NATIONAL ASSESSMENTS? (Enter an integer) ___

Q7 How many PROFESSIONAL DEVELOPMENT SESSIONS (e.g., school or district workshop, state or national conference workshop) have you completed on CLASSROOM AND/OR NATIONAL ASSESSMENTS? (Enter integer) ___

Q8 How often do your PLC's meet at your school?

- ❍ Daily
- ❍ Weekly
- ❍ Bi-Weekly
- ❍ Monthly
- ❍ Quarterly

Section 2: Assessment Beliefs

For the second set of questions, rate the degree to which you agree with the following statements as they apply to your current role as administrator or PLC facilitator at your school.

I believe that STUDENT ASSESSMENT DATA are the primary source for…	Strongly Disagree	Disagree	Agree	Strongly Agree
Q1...understanding what students KNOW.	□	□	□	□
Q2...understanding what students CAN DO.	□	□	□	□
Q3…making assessment decisions about what TYPES of assessments to use.	□	□	□	□
Q4…making decisions about WHICH TEACHERS are assigned to teach certain content areas.	□	□	□	□
Q5…making curriculum decisions about WHAT to teach.	□	□	□	□
Q6…making instructional decisions about HOW to teach.	□	□	□	□

I believe that CLASSROOM CONTEXT (e.g., physical settings, student groups, learning activities or tasks, transitions) is as important as content when...	Strongly Disagree	Disagree	Agree	Strongly Agree
Q7...deciding WHAT to teach.	☐	☐	☐	☐
Q8...deciding HOW to teach.	☐	☐	☐	☐
Q9...assessing student learning outcomes.	☐	☐	☐	☐

I believe that STUDENT CONTEXT (e.g., prior knowledge, experiences, motivations, attitudes, learning styles) is as important as content when...	Strongly Disagree	Disagree	Agree	Strongly Agree
Q10...deciding WHAT to teach.	☐	☐	☐	☐
Q11...deciding HOW to teach.	☐	☐	☐	☐
Q12…assessing student learning outcomes.	☐	☐	☐	☐

I believe that…	Strongly Disagree	Disagree	Agree	Strongly Agree
Q13…teachers analyzing multiple forms of student assessment data is as important as teaching.	☐	☐	☐	☐
Q14...leaders promoting effective assessment practices is as important as promoting teaching practices.	☐	☐	☐	☐

Section 3: Assessment Practices

For the third set of questions, indicate how often you engage in these activities as administrator or PLC facilitator at your current school.

Q1 I discuss with teachers the school's vision for student assessments.

- ❍ Almost Never
- ❍ Seldom
- ❍ Occasionally
- ❍ Frequently
- ❍ Almost Always

Q2 I engage teachers in reflective dialogue about student assessments.

- ❍ Almost Never
- ❍ Seldom
- ❍ Occasionally
- ❍ Frequently
- ❍ Almost Always

Q3 I initiate instructional planning with teachers based on student assessment data.

- ❍ Almost Never
- ❍ Seldom
- ❍ Occasionally
- ❍ Frequently
- ❍ Almost Always

Q4 I reflect on my own student assessment knowledge.

- ❍ Almost Never
- ❍ Seldom
- ❍ Occasionally
- ❍ Frequently
- ❍ Almost Always

I provide guidance to teachers on...	Almost Never	Seldom	Occasionally	Frequently	Almost Always
Q5...writing assessment items.	□	□	□	□	□
Q6...scoring assessment items.	□	□	□	□	□
Q7...organizing assessment data into charts and graphs.	□	□	□	□	□
Q8...analyzing assessment data trends.	□	□	□	□	□
Q9...using student assessment data to adjust instruction.	□	□	□	□	□
Q10...using student assessment data to change assessment items.	□	□	□	□	□

As an assessment leader, I set expectations that teachers...	Almost Never	Seldom	Occasionally	Frequently	Almost Always
Q11...identify student-learning targets	☐	☐	☐	☐	☐
Q12...develop formative classroom assessments aligned with student-learning targets.	☐	☐	☐	☐	☐
Q13...design summative classroom assessments aligned with student-learning targets.	☐	☐	☐	☐	☐
Q14...collect multiple forms of student-assessment data.	☐	☐	☐	☐	☐
Q15...analyze multiple forms of student-assessment data.	☐	☐	☐	☐	☐
Q16…modify curriculum based on outcomes from student-assessment data.	☐	☐	☐	☐	☐
Q17…adjust instruction based on outcomes from student-assessment data.	☐	☐	☐	☐	☐
Q18…select assessment tools based on outcomes from student-assessment data.	☐	☐	☐	☐	☐

Section 4: Assessment Knowledge

For the next 5 sets of questions, read each scenario and respond to the questions related to each scenario.

Scenario 1: Ms. O'Connor, a mathematics teacher, questions how well her 10th grade students can apply what they have learned in class to everyday life situations. Although the teacher's manual contains numerous test items, she is not convinced that giving a paper-and-pencil test is the best method for determining what she wants to know.

Q1.1 Based on the above scenario, the type of assessment that best answers Ms. O'Connor's question is called a/an

- performance assessment
- authentic assessment
- extended response assessment
- standardized test

Q1.2 In order to grade her students' knowledge accurately and consistently, Ms. O'Connor is advised to

- identify criteria from the unit objectives and create a scoring rubric.
- develop a scoring rubric after getting a feel for what students can do.
- consider student performance on similar types of assignments.
- consult with experienced colleagues about criteria that has been used in the past.

Q1.3 To determine how well her students perform in mathematics compared to other 10th graders, Ms. O'Connor administers a standardized mathematics test. This practice is acceptable ONLY if the

- reliability of the standardized test does not exceed .60.
- standardized test is administered individually to students.
- content of the standardized test is well known to students.
- comparison group is comprised of grade level peers.

Q1.4 Which of the following is NOT an appropriate use of the results from this standardized mathematics test?

- planning instruction
- assigning student grades
- determining students' strengths and weaknesses
- developing curriculum

Q1.5 Throughout instruction, Ms. O'Connor assesses how well her students grasp the material. These assessments range from short quizzes following the introduction of a new topic to administering an end-of -unit final exam. In order to improve the validity of this grading procedure, Ms. O'Connor should

- make the grading scale the same for all assessments.
- consider students' prior performance before assigning a final grade.
- weight assessments according to their relative importance.
- take into consideration each student's effort when calculating grades.

Q1.6 During a parent-teacher conference, one of the parents asks Ms. O'Connor what it means that his daughter scored in the 80th percentile in mathematics. Which of the following provides the best explanation of this student's score?

- She got 80% of the items on the mathematics test correct.
- She is likely to earn a grade of 'B' in her mathematics class.
- She is demonstrating above grade level performance in mathematics.
- She scored the same or better than 80% of the norm group.

Q2.7 Which of the following is an appropriate use of assessment information?

- Utilize information from a variety of assessments when making decisions about student learning.
- Use scores from standardized tests to determine teacher instructional effectiveness.
- Use scores from a standardized test as the primary indicator of student retention.
- Post final grades in order to provide normative information to students in the class.

Scenario 2: Mr. Okawa, a fifth-grade teacher, is planning instruction for the next grading period, aware of the fact that his students will be taking the statewide achievement test near the end of the grading period.

Q2.1 Mr. Okawa's mathematics unit for this grading period focuses on multi-step problem-solving. He wants to assess students' problem-solving abilities at the end of the unit to determine if any re-instruction is necessary prior to the statewide test. Which of the following assessment strategies are the most appropriate choice? He should choose

- the assessment included in the teacher's manual from the textbook he uses.
- an assessment which is consistent with the content and skills he taught.
- a different standardized assessment that provides a score on similar skills.
- an assessment which covers single-step problem-solving skills.

Q2.2 Mr. Okawa decides to develop his own assessment in order to determine if reinstruction is necessary. He also wants to use his assessment as a means of anticipating how his students will perform on the statewide assessment. In order for him to accurately approximate his students' performance, which of the following would be the most appropriate type of assessment for him to develop?

- a performance assessment
- a multiple-choice test
- a portfolio assessment
- an essay test

Q2.3 Julie, one of Mr. Okawa's students, received a percentile rank of 60 on the problem-solving skills subtest of the statewide assessment. This score is most appropriately interpreted as which of the following?

- ❍ Above average
- ❍ Below average
- ❍ At the national average
- ❍ Not enough information

Q2.4 Juan, another student in Mr. Okawa's class, received a scaled score of 196 on the reading comprehension subtest of the statewide assessment. The cut score is 200; therefore, Juan does not pass the subtest. However, the subtest has a standard error of measurement equal to 6. Which of the following is the best decision for Mr. Okawa to make regarding appropriate instruction to meet Juan's needs?

- ❍ Juan has clearly not achieved the minimum level of reading comprehension and should receive remedial reading instruction.
- ❍ Mr. Okawa knows that Juan could have scored higher, so the results of the test should be ignored.
- ❍ Juan may likely have achieved the minimum level of reading comprehension and nothing different or additional should be done.
- ❍ Mr. Okawa knows that Juan should have scored much lower, so the results of the test should be ignored.

Q2.5 Which of Mr. Okawa's grading practices is LEAST reflective of achievement?

- ❍ Daily homework and chapter tests
- ❍ Daily homework and chapter tests, with points deducted for poor effort
- ❍ Daily homework and chapter tests, where students are permitted to redo assignments in order to meet higher standards
- ❍ Chapter tests, where daily homework is not formally graded

Q2.6 Barbara scored at the 60th percentile on mathematics problem-solving skills and at the 56th percentile on reading comprehension. The percentile bands for each test are five percentile ranks wide. What advice should Mr. Okawa give to Barbara's parents? They should

- ❍ ignore the difference; her performance was essentially the same on the two tests.
- ❍ seek additional tutoring help for Barbara in reading.
- ❍ force Barbara to read more at home.
- ❍ provide enrichment experiences for Barbara in math, which is her better performance area.

Q2.7 Mr. Okawa is worried that his students would not perform well on the statewide assessment. He did all of the following to help increase students' scores. Which was unethical?

- He instructed students in strategies for taking multiple-choice tests, such as how to use answer sheets.
- He planned his instruction so that it focused on concepts and skills to be covered on the test.
- He encouraged the students to do their best, and provided them with a reward after testing was complete.
- He allowed students to practice with items from an alternate form of the test.

Scenario 3: Ms. Green is an eighth-grade American History teacher. She just finished teaching a unit on the Industrial Revolution and wants to assess her students' higher-order thinking skills. Ms. Green decided to give her students a single assessment in the form of an end-of-unit multiple-choice test. She anticipates that most of her students will perform well on the test.

Q3.1 Based on her goal, what can you conclude about her decision to administer a multiple-choice test?

- This is an appropriate choice for a unit assessment.
- The test scores may not be valid for this purpose.
- The test scores may not be reliable for this purpose.
- A true-false test would be more appropriate.

Q3.2 To determine the quality of her multiple-choice test, Ms. Green conducts an item analysis and examines all of the following EXCEPT

- item difficulty values.
- item discrimination values.
- reliability coefficients.
- validity coefficients.

Q3.3 Ms. Green decides to score the tests using a 100-percent correct scale. Generally speaking, what is the proper interpretation of a student score of 85 on this scale? The student

- answered 85% of the items on the test correctly.
- knows 85% of the content covered by this instructional unit.
- scored higher than 85% of other students who took this test.
- scored lower than 85% of other students who took this test.

Q3.4 Some of Ms. Green's students do not score well on the multiple-choice test. She decides that the next time she teaches this unit, she will begin by administering a pretest

to check for students' prerequisite knowledge. She will then adjust her instruction based on the pretest results. What type of information is Ms. Green using?

- ❍ norm-referenced information
- ❍ criterion-referenced information
- ❍ both norm- and criterion-referenced information
- ❍ neither norm- nor criterion-referenced information

Q3.5 The Industrial Revolution test is the only student work that Ms. Green grades for the current grading period. Therefore, grades are assigned only on the basis of the test. What is the major criticism of this practice?

- ❍ The test, and therefore the grades, reflect too narrow a curricular focus.
- ❍ These grades, since based on tests alone, are probably biased against some minority students.
- ❍ She should add extra points to the scores of students who scored low on the test.
- ❍ Decisions like grades should be based on more than one piece of information.

Q3.6 Mr. Simpson, another American History teacher, bases his grades primarily on his observations of students during class. The primary distinction between his system of assigning grades and that used by Ms. Green is BEST characterized as which of the following?

- ❍ Ms. Green uses formal assessment; Mr. Simpson uses informal assessment.
- ❍ Ms. Green uses formative assessment; Mr. Simpson uses summative assessment.
- ❍ Ms. Green uses standardized assessment; Mr. Simpson uses nonstandardized assessment.
- ❍ Ms. Green uses traditional assessment; Mr. Simpson uses alternative assessment.

Q3.7 Based on their grades from last year, Ms. Green believes that some of her low-scoring students are brighter than their test scores indicate. Based on this knowledge, she decides to add some points to their test scores, thus raising their grades. Which of Ms. Green's actions was unethical?

- ❍ examining her students' previous academic performance
- ❍ adjusting grades in her course
- ❍ using previous grades to adjust current grades
- ❍ adjusting some students' grades and not others'

Scenario 4: Mr. Valdez is an English teacher in the newly built middle school. Experienced in issues of classroom assessment, Mr. Valdez is often asked to respond to the district's questions concerning best practices for evaluating student learning.

Q4.1 Ms. Franklin, also an English teacher, asks what type of assessment is best for evaluating her 6th graders' writing skills. Which of the following methods is likely the BEST response to her question?

- selected response methods
- true/false statements
- completion items
- essay prompts

Q4.2 One of the middle school mathematics teachers is redesigning her tests to make greater use of "story problems" as a way to check students' mathematics understanding. She consults with Mr. Valdez to see what, if any, concerns she should be aware of when constructing assessments of this type. Which statement is not an appropriate recommendation when designing story-based mathematics tests?

- make sure that the reading level is grade appropriate
- avoid scenarios more familiar to certain groups over others
- check for clarity of sentence construction
- incorporate scenarios used during instruction

Q4.3 Isabel, a student in Mr. Valdez's class, scored 78 points on a standardized English test which had a mean of 80 and a standard deviation of 4. She scored 60 points on the science portion of this test which had a mean of 50 and a standard deviation of 3. Based on the above information, in comparison to her peers, which statement provides the most accurate interpretation?

- Isabel is better in English than in science.
- Isabel is better in science than in English.
- Isabel is below average in both subjects.
- Isabel is close to average in both subjects.

Q4.4 At the end of each class period, Mr. Valdez does a quick "check in" with his students to get an impression of their understanding. In this example, the primary purpose for conducting formative assessment is to

- identify cumulative knowledge.
- determine content for the final exam.
- plan classroom instruction.
- evaluate curriculum appropriateness.

Q4.5 To prepare students for state testing and identify areas of school improvement, all 6th grade English teachers give a common final exam which contains a series of essay items. Recently, however, several teachers expressed concern that the time and effort necessary to complete grading may result in inconsistent scoring. They consult with Mr.

Valdez. Which of the following provides the BEST response to the teachers' concerns for consistency?

- ❍ grade all responses to essay #1 before grading responses to essay #2
- ❍ during grading, adjust rubric criteria to reflect exemplary student work
- ❍ utilize a holistic scoring method to minimize teacher subjectivity in scoring
- ❍ all things being equal, it is best to limit the use of multiple essay exams

Q4.6 Jeremy, a 6th grade student in Mr. Valdez's class, received a grade equivalent score of 7.2 on a standardized reading test. Jeremy's parents wonder what this means. Based on the above information, which of the following statements provides the most appropriate interpretation of this student's score?

- ❍ Jeremy is reading at the 7th grade level.
- ❍ Jeremy is reading better than the majority of students in his class.
- ❍ Jeremy is reading 6th grade material as expected.
- ❍ Jeremy should be placed in a 7th grade reading class.

Q4.7 "To ensure that standardized test results provide an accurate picture of what students really know, it is recommended that teachers clarify items that are confusing to students." Based on best practices of assessment, which of the following is an appropriate response to the above statement?

- ❍ This is an acceptable way to reduce error in testing.
- ❍ This is an acceptable way to increase test validity.
- ❍ This is unacceptable because it labels students as poor readers.
- ❍ This is unacceptable because it breaks standardization.

Scenario 5: Ms. Hawkins is responsible for teaching science at the 4th grade level. Over the past couple of years, her students have struggled with investigations of how water changes from one state to another (i.e., freezing, melting, condensing, and evaporating), but she is unsure of where the specific difficulties lie. She is aware that her students need to improve their conceptual understanding of this content standard.

Q5.1 Ms. Hawkins wishes to conduct an assessment in order to identify the specific difficulties her students are experiencing. Which of the following would best meet her needs?

- ❍ diagnostic assessment
- ❍ informal assessment
- ❍ standardized assessment
- ❍ summative assessment

Q5.2 In an effort to refine both her instruction and assessment of this content, Ms. Hawkins conducts an item analysis of student scores from last year's final unit test over this material. She should definitely discard or substantially revise a test item that has a

- ❍ difficulty value between .50 and .75.
- ❍ discrimination value equal to +.30.
- ❍ discrimination value equal to -.50.
- ❍ difficulty value equal to .90.

Q5.3 Ms. Hawkins' unit test also includes a restricted-response essay item. She is concerned with the demonstrated level of understanding of several specific criteria in her students' responses. Which of the following would best facilitate her scoring of these responses?

- ❍ objective answer key
- ❍ holistic rubric
- ❍ checklist
- ❍ analytic rubric

Q5.4 Following the completion of the unit, Ms. Hawkins determines that her students have satisfactorily mastered these concepts. However, when her students take the statewide standardized assessment in the spring, she notices that her students perform very poorly on items addressing these same concepts. Considering the discrepancy between students' classroom performance and their standardized test results, what action is most appropriate when making decisions concerning school improvement?

- ❍ recommend that classroom instruction be consistent among 4th grade science teachers
- ❍ ensure alignment between instruction and what is measured on the standardized test
- ❍ select a standardized test that is more likely to yield higher scores in science
- ❍ identify the percentage of students predicted to perform well in advanced science classes

Q5.5 Ms. Hawkins wants to be sure that the term grades she assigns to her students' performance in science reflect each student's respective level of content mastery for that unit. Which of the following grading systems would best accomplish this goal?

- ❍ criterion-referenced grading system
- ❍ norm-referenced grading system
- ❍ pass–fail grading system
- ❍ portfolio grading system

Q5.6 Nolan is a student in Ms. Hawkins' class. He receives a raw score of 12 items answered correctly out of a possible 15 on the physical science portion of a standardized test. This raw score equates to a percentile rank of 45. His parents are confused about how he could answer so many items correctly, but receive such a low percentile rank.

They approach Ms. Hawkins for a possible explanation. Which of the following is the appropriate explanation to offer to his parents?

- ❍ "I don't know…there must be something wrong with the way the test company figured the scores."
- ❍ "Although Nolan answered 12 correctly, numerous students answered more than 12 correctly."
- ❍ "Raw scores are purely criterion-referenced and percentile ranks are merely one form of norm referenced scoring."
- ❍ "Raw scores are purely norm-referenced and percentile ranks are merely one form of criterion referenced scoring."

Q5.7 In an attempt to encourage and motivate her students who are struggling academically, Ms. Hawkins shares her gradebook, especially test scores, with them in order to demonstrate how well others are performing. Another teacher advises her not to do this, as it is a clear violation of

- ❍ The Code of Fair Testing Practices in Education.

The Family and Education Rights and Privacy Act.

- ❍ The Standards for Teacher Competence in the Educational Assessment of Students.
- ❍ The No Child Left Behind Act.

Section 5: Demographics

For the last set of questions, provide your best responses.

Q1 What is your highest level of education?

- ❍ Bachelor
- ❍ Master
- ❍ Specialist
- ❍ Doctorate

Q2 What is your age in years (e.g., 45)?

Q3 What is your ethical background?

- ❍ White, not Hispanic
- ❍ Black or African American
- ❍ Hispanic or Latino
- ❍ American Indian and Alaska Native
- ❍ Asian
- ❍ Native Hawaiian and Pacific Islander
- ❍ Other

Q4 Which of the following best describes you?

- ❍ Female
- ❍ Male

Q5. Assessment leadership is defined in the literature as instructional leaders who (1) establish a vision that sets clear and appropriate expectations for student assessment systems, (2) lead data discussions, (3) foster assessment literacy in teachers through ongoing, collaborative learning experiences, and (4) self-reflect on personal assessment practices. To what extent does this definition match your current role? Would you rewrite the definition? If so, what would it be?

REFERENCES

Acock, A. (2005). Working with missing values. *Journal of Marriage and Family, 67*(4), 1012-1028.

Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education, 39*(2), 113-123.

Allison, P. (2001). *Missing data.* Thousand Oaks, CA: Sage

American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990). *Standards for Teacher Competence in Educational Assessment of Students*. Retrieved from http://buros.org/standards-teacher-competence-educational-assessment-students

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice, 48*, 12-19.

Arter, J. A., Stiggins, R. J., Duke, D., & Sagor, R. (1993). Promoting assessment literacy among principals. *NASSP Bulletin, 77*(556), 1–7.

Bernhardt, V. (2004). *Data analysis for continuous school improvement.* Larchmont, NY: Eye on Education.

Beziat, T. L., & Coleman, B.K. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher, 27*(1), 25-30.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-71.

Blink, R. (2007). *Data-driven instructional leadership.* Larchmont, NY: Eye on Education.

Boudett, K. P., City, E. A., & Murnane, R. J. (2010). *Datawise: A step by step guide to using assessment results to improve learning and teaching.* Cambridge, MA: Harvard Education Press.

Brick, J., & Kalton, J. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research, 5*, 215-238.

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education, 8*(2), 153-169.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Assessment Issues and Practice 30*(1), 3-12.

Brown, T.A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Bryan, M. & Simone, S. (2012). *2010 College Course Map* (NCES 2012-162REV). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, & U.S. Department of Education. Retrieved from http://nces.ed.gov/pubsearch

Buck, G. A., & Trauth-Nare, A. E. (2009). Preparing teachers to make the formative assessment process integral to science teaching and learning. *Journal of Science Teacher Education, 20*(5), 475-494.

Çalik, T., Sezgin, F., Kavgaci, H., & Kilinç, A. C. (2012). Examination of relationships between instructional leadership of school principals and self-efficacy of teachers and collective teacher efficacy. *Kuram Ve Uygulamada Egitim Bilimleri, 12*(4), 2498-2504.

Campbell, C., Murphy, J. A., & Holt, J. K. (2002). *Psychometric analysis of an*

assessment literacy instrument: Applicability to preservice teachers. Columbus, OH: Mid-Western Educational Research Association.

Carr, D. L. (2002). *An analysis of selected elementary principals' instructional leadership practices and uses of student assessment information in the relation to the categorical ratings of schools in North Carolina* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3345788).

Cheema, J.R. (2014). Some general guidelines for choosing missing data handling methods in educational research. *Journal of Modern Applied Statistical Methods, 13*(2), 53-75.

Chen F. F., Hayes A., Carver C. S., Laurenceau J. P., & Zhang Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80,* 219-251.

Clark, I. (2011). Formative assessment: Policy, perspectives and practice. *Florida Journal of Educational Administration & Policy, 4*(2), 158-180.

Clifford, M., & Mason, C. (2013). *Leadership for the Common Core: More than one thousand school principals respond.* Alexandria, VA: National Association of Elementary School Principals. Retrieved from http://www.afterschoolalliance.org/issue_briefs/issue_common_core_63.pdf

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*(4), 469-495.

Collins, J. (2001). *Good to great.* New York, NY: Harper Collins.

Collinson, V., & Cook, T. (2007). *Organizational learning: Improving learning, teaching and leading in school systems.* Thousand Oaks, CA: Sage Publications, Inc.

Coggshall, J. G., Bivona, L., & Reschly, D. J, (2012, August). *Evaluating the effectiveness of teacher preparation programs for support and accountability* [Research and policy brief]. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/TQ_RandP_BriefEvaluatingEffectiveness.pdf

Copeland, M. (2000). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis, 25*(4), 375-395.

Cotton, K. (2003). *Principals and student achievement: What the research says.* Alexandria, VA: Association for Supervision and Curriculum Development

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles, CA: Sage.

Danielson, C. (2009). *Talk about teaching: Leading professional conversations*. Thousand Oaks, CA: Corwin Press.

Darling-Hammond, L. (2004). Standards, accountability, and school reform, *Teachers College Record, 106*(6), 1047-1085.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment, 21*(4), 248-266.

Deneen, C. C., & Brown, G. T. (2060). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education, 3*(1), 1-14.

Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy, 16*, 731-762.

Dong, Y., & Peng, C-Y.J. (2013). Principled missing data methods for researchers. *Springer Plus, 2,* 1-17. Retrieved from https://springerplus.springeropen.com/articles/10.1186/2193-1801-2-222

DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement.* Bloomington, IN: Solution Tree.

Dufresne, P., & McKenzie, A. S. (2009). A culture of ethical leadership. *Principal Leadership, 10*(2), 36-39.

Dunn, J. L., & Allen, J. (2009). Holding schools accountable for the growth of nonproficient students: Coordinating measurement and assessment. *Educational Measurement: Issues and Practice, 28*(4), 27–41.

Earl, L., & Fullan, M. (2003). Using data in leadership for learning. *Cambridge Journal of Education, 33*(3), 383-394.

Editorial Projects in Education Research Center. (2016, March 31). The Every Student Succeeds Act: An ESSA overview. *Education Week.* Retrieved from http://www.edweek.org/ew/issues/every-student-succeeds-act/

Elfers, A. M. (2008). Teachers' response to standards-based reform: Probing reform assumptions in Washington state. *Education Policy Analysis Archives, 16*(8), 1-29.

Enders, C.K. (2003). Using the expectation maximization algorithm to estimate coefficient coefficient alpha for scales with item-level missing data. *Psychological Methods, 8*(3), 322–337.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: The Guilford Press.

Fisher, D. (2005). The missing link: Standards, assessment, and instruction. *Voices from the Middle, 13*(2), 8-11.

Frey, B. B., & Schmitt, V. L. (2010). Teachers' classroom assessment practices. *Middle Grades Research Journal, 5*(3), 107-117.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco, CA: Jossey-Bass.

Gallagher, L., Means, B., & Padilla, C. (2008). *Teachers' use of student data systems to improve instruction: 2005 to 2007* [Report prepared for U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service]. Menlo Park, CA: SRI International.

Gedik, S., & Bellibas, M. S. (2015). Examining schools' distributed instructional leadership capacity: Comparison of elementary and secondary schools. *Journal of Education and Training Studies, 3*(6), 101-110.

Goertz, M. E., Olah, L. *n*., & Riggan, M. (2009). *From testing to teaching: The use of interim assessments in classroom instruction.* Philadelphia, PA: Consortium for Policy Research in Education. Retrieved from http://files.eric.ed.gov/fulltext/ED519791.pdf

Gorsuch, R.L. (1988). Exploratory factor analysis. In J. R. Nesselroade & Cattell, R. B. (Eds.), *Handbook of multivariate experimental psychology* (pp. 231-258). New York, NY: Plenum Press.

Green, R. L. (2011). *The four dimensions of principal leadership: A framework for leading 21st century schools*. Boston, MA: Allyn & Bacon.

Greenberg, J., McKee, A., & Walsh, K. *(*2013*). Teacher prep review: A review of the nation's teacher preparation programs.* Washington, D C: National Council on Teacher Quality. Retrieved from http://www.nctq.org/dmsView/Teacher_Prep_Review_executive_summary

Guskey, T. R. (2003). How classroom assessments improve student learning. *Educational Leadership, 60*(5), 6-11.

Hall, B.W. & Hewitt-Gervais, C.M. (2000). The application of student portfolios in primary-intermediate and self- contained-multiage team classroom environments: Implications for instruction, learning, and assessment. *Applied Measurement in Education, 13*(2), 209-228.

Hallinger, P. (2001, April). *A review of two decades of research of the principalship using the Principal Instructional Management Rating Scale.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Hallinger, P. (2011). A review of three decades of doctoral studies using the Principal Instructional Management Rating Scale: A lens on methodological progress in educational leadership. *Educational Administration Quarterly, 47*(2), 271-306.

Hallinger, P., & Heck, R. H. (1996). Reassessing the principal‘s role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly, 32*(1), 5-44.

Hallinger, P., & Heck, R. H. (2010). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School Leadership & Management, 30*(2), 95-110.

Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The new instructional leadership: Creating data-driven instructional systems in school. *Journal of School Leadership, 17*(2), 159-194.

Hameister, T. (2013). *Assessment leadership to promote student learning* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3551293).

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management, 24(2*), 297-327.

Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.

Henry, S. S. (2011). *Principals' use of assessment data to drive student academic achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3489142)

Horng, E., & Loeb, S. (2010). New thinking about instructional leadership. *Phi Delta Kappan, (92)*3. Retrieved March 31, 2012, from http://web.ebscohost.com.ezproxy.liberty.edu:2048/ehost/pdfviewer/pdfviewer?vid=3&hid=10&sid=c456f1a7-8f77-48e3-8639-66ba53a7134a%40sessionmgr10

Hoy, W. K., & Miskel C. G. (2008). *Educational administration: Theory, research, and practice* (8th ed.). New York, NY: McGraw Hill.

Hu, L. T., & Bentler, P. M. (2009). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453.

Hursh, D. (2007). Exacerbating inequality: The failed promise of the No Child Left Behind Act. *Race, Ethnicity, and Education, 10*(3), 295-308.

Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models.* Thousand Oaks, CA: Sage.

Impara, J. C., & Plake, B. S. (1995). Comparing counselors', school administrators', and teachers' knowledge in student assessment. *Measurement and Evaluation in Counseling and Development, 28*, 78-87.

Impara, J. C., Plake, B. S., & Merwin, J. (1994). Student assessment tasks and knowledge: Comparing superintendents and elementary and secondary principals. *Journal of School Leadership, 4*(5), 517–529

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258–1287.

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis; An overview and some recommendations. *Psychological Methods, 14*(1), 6-23.

Jenkins, B. (2009). What it takes to be an instructional leader. *Principal, 88*(3), 34-37.

Johnson, D.R., & Young, R. (2011). Toward best practices in analyzing datasets with missing data: Comparisons and recommendations. *Journal of Marriage and Family, 73*(5), 926-945.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31–36.

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *The Journal of Economic Perspectives, 16* (4), 91-114.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*(4), 496-520.

King, S. P., & Amon, C. (2008). Assessment data: A tool for student and teacher growth. In E. D. Mandinach & M. Honey (Eds.), *Data-driven school improvement: Linking data and learning* (pp. 71–86). New York: Teachers College Press.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28-37.

Knapp, M. S., Swinnerton, J. A., Copland, M. A., & Monpas, J. (2006). *Data-informed leadership in education.* Seattle: University of Washington, Center for the Study of Teaching and Policy.

Langkamp, D. L., Lehman, A., & Lemeshow, S. (2010). Techniques for handling missing data in secondary analyses of large surveys. *Academic Pediatrics*, *10*(3), 205-210.

Lashway, L. (2003, July). *Role of the school leader: Trends and issues*. Eugene, OR: Clearinghouse on Educational Management. Retrieved from http://www.schoolsmovingup.net/saelp/rs/role_of_the_school_leader.pdf

Leithwood, K. (2008). Should educational leadership focus on best practices or next practices? *Journal of Educational Change*, *9*(1), 71-75.

Leithwood, K., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership and Management*, *28*(1), 27-42.

Leithwood, K., & Jantizi, D. (2005). A review of transformational school leadership research, 1996-2005. *Leadership and Policy in Schools, 4*(3), 177-199.

Leithwood, L., & Louis, K. (2012). *Linking leadership to student learning*. San Francisco, CA: Jossey-Bass.

Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student achievement*. Minneapolis, MN: Center for Applied Research and Educational Improvement and Ontario Institute for Studies in Education.

Leithwood, K., & Mascall, B. (2008). Collective leadership effects on student achievement. *Educational Administration Quarterly, 44*(4), 529-561.

Linn, R. L. (2000). Assessments and accountability. *Education Researcher*, *29*(2), 4–15. Available from http://www.aera.net/pubs/er/arts/29-02/linn01.htm

Lie, Y., & De, A. (2015). Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *International Journal of Statistical Medical Research, 4*(3), 287-295.

Loeb, H. L., Knapp, M. S., & Elfers, A. M. (2008). Teachers' response to standards-based reform: Probing reform assumptions in Washington State. *Education Policy Analysis Archives*, *16*(8). Retrieved from http://epaa.asu.edu/epaa/v16n8/

Louis, K. S., Leithwood, K., Wahlstrom, K. L., & Anderson, S. E. (2010). *Learning from leadership: Investigating the links to improved student learning.* New York, NY: The Wallace Foundation.

Manly, C. A., & Wells, R. S. (2015). Reporting the use of multiple imputation for missing data in high education research. *Research in Higher Education, 56*, 397-409.

Marks, H. M., & Nance, J. P. (2007). Contexts of accountability under systemic reform: Implications for principal influence on instruction and supervision. *Educational Administration Quarterly, 43*(1), 3-37.

Marsh, J., Pane, J., & Hamilton, L. (2006). *Making sense of data-driven decision making in education*. [Occasional paper.] Santa Monica, CA: RAND Corporation.

Marzano, R. J. (2006). *Classroom assessment & grading that work.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R.J. (2010). The art and science of teaching: When students track their progress. *Health and Learning, 67*(4), 86-87.

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works - from research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.

Matthews, K. W. (2007). *The relationships among middle school teachers' assessment practices, instructional leadership, and student achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3322724).

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd ed). Thousand Oaks, CA: Sage.

McMillan, J. H. & Schumacher, S. (2010). *Research in education: Evidence-based inquiry* (7th ed.). Boston, MA: Pearson.

McShane, M. Q., Lautzenheiser, D., Deane, K. C., & Kimmel, A. (2012). What's ahead for education after the 2012 election. *Education Outlook, 8.* Retrieved from http://www.aei.org/files/2012/12/03/-whats-ahead-for-education-after-the-2012-election_171852704398.pdf

Means, B., Padilla, C., & Gallagher, L. (2010). *Use of education data at the local level: From accountability to instructional improvement* [Report prepared for U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service]. Menlo Park, CA: SRI International. Retrieved from https://www2.ed.gov/rschstat/eval/tech/use-of-education-data/use-of-education-data.pdf

Mertler, C. A. (2003, October). *Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH. Retrieved from http://wps.ablongman.com/wps/media/objects/3984/4080433/tchassesslit.pdf

Mertler, C. A. (2005). The role of classroom experience in preservice and inservice teachers' assessment literacy. *Mid-Western Educational Researcher, 18*(4), 25-33.

Mertler, C. A., & Campbell, C. C. (2005, April). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the Assessment Literacy Inventory*. Paper presented at the annual meeting of the American Educational Research Association, Quebec, CA.

Meyers, C. V., & Murphy, J. (2007). Turning around failing schools: An analysis. *Journal of School Leadership, 17*, 651-639.

Militello, M., Schweid, J., & Sireci, S. G. (2010). Formative assessment systems: Evaluating the fit between school districts' needs and assessment systems' characteristics. *Educational Assessment, Evaluation, and Accountability, 22*(1), 29-52.

Mintrop, H. (2012). Bridging accountability obligations, professional values, and (perceived) student needs with integrity. *Journal of Educational Administration, 50*(5), 695-726.

Muñoz, M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education. *Phi Delta Kappan, 96*(7), 64.

Muthén, L. K., & Muthén, B. O. (1998–2012). Mplus user's guide (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Governors Association. (2010). *Common core state standards*. Washington, DC: Author. Retrieved from: http://www.corestandards.org/the-standards

Noonan, B., & Renihan, P. (2008). Building assessment literacy: Perceptions of high school principals. *International Electronic Journal for Leadership in Learning, 56*(4), 1-21.

Noonan, B., & Renihan, P. (2006). Demystifying assessment leadership. *Canadian Journal of Educational Administration and Policy, 56*(4), 1-20.

O'Day, J. (2002). Complexity, accountability, and school improvement. *Harvard Educational review, 72*(3), 293-329.

O‘Donnell, R. J., & White, G. P. (2005). Within the accountability era: Principals’ instructional leadership behaviors and student achievement. *NASSP Bulletin, 89*(645), 56–71.

Pampaka, M., Hutcheson, G., & Williams, J. (2014). Handling missing data: Analysis of a challenging data set using multiple imputation. *International Journal of Research and Method in Education, 39*(1), 19-37.

Penuel, W. R., Tatar, D. G., & Roschelle, J. (2004). The role of research on contexts of teaching practice in informing the design of handheld learning technologies. *Journal of Educational Computing Research, 30*(4), 353–370.

Perry, M. L. (2013). *Teacher and principal assessment literacy* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3568118).

Peng, C., Harwell, M., Liou, S., & Ehman, L. (2006). Advances in missing data methods and implications for educational research. In S. S. Sawilowsky (Ed.), *Real Data Analysis* (pp. 31-78). Charlotte, NC: New Information Age.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research. 74*(4), 525–556.

Plake, B. S. (1993). Teacher assessment literacy: Teachers’ competencies in the educational assessment of students. *Mid-Western Educational Researcher, 6*(1), 21-27.

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement. Issues and Practice, 12*(4), 10-12, 39.

Popham. W. J. (2008). *Everything school leaders need to know about assessment*. Thousand Oaks, CA: Sage.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice, 48*(1), 4-11.

Preacher, K. J., Cai, L., & MacCallum, R. C. (2007). Alternatives to traditional model comparison strategies for covariance structure models. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 33-62). Mahwah, NJ: Lawrence Erlbaum Associates.

Preuss, P. (2007). *Data-driven decision making and dynamic planning: A school leader's guide.* Larchmont, NY: Eye on Education.

Prytula, M., Noonan, B., & Hellsten, L. (2013, March). Toward instructional leadership: Principals' perceptions of large-scale assessment in schools. *Canadian Journal of Educational Administration and Policy*, 140, 1-30.

Renihan, P., & Noonan, B. (2012). Principals as assessment leaders in rural schools. *Rural Educator, 33*(3), 1-8.

Robinson, V. J., Lloyd, C., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership type. *Educational Administration Quarterly, 44*(5), 635-674.

Roschelle, J., Penuel, W.R., Yarnall, L., Shechtman, N., & Tatar, D. (2005). Handheld tools that 'informate' assessment of student learning in science: A requirements analysis. *Journal of Computer Assisted Learning, 21*(3), 190–203.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.

Sanders, N. M., & Kearney, K. M. (2008). *Performance expectations and indicators for education leaders: An ISLLC-based guide to implementing leader standards and a companion guide to the educational leadership policy standards ISLLC 2008.* Washington, DC: Council of Chief State School Officers. Retrieved from http://www.ccsso.org/Documents/2008/Peformance_Indicators_2008.pdf

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London, UK: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177.

Sebastian, J., & Allensworth, E. (2012). The influence of principal leadership on classroom instruction and student learning: A study of mediated pathways to learning. *Educational Administration Quarterly, 48*(4), 626-663.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M.K. & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education, 21*(4), 295-314.

Shen, J., Cooley, V. E., Reeves, P., Burt, W. L., Ryan, L., Rainey, J. M., & Yuan, W. (2010). Using data for decision making: Perspectives from 16 principals in Michigan, USA. *International Review of Education, 56*(4), 435-456.

Sterrett, W. L. (2005). *Perceived instructional leadership role in elementary principals in Virginia* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3169662).

Stiggins, R. (2001). *Student involved classroom assessment* (3rd ed.). Columbus, OH: Merrill Prentice Hall.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324–328.

Stiggins, R., & Duke, D. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan, 90*(4), 285-291.

Supovitz, J. A., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement*. Philadelphia, PA: Consortium for Policy Research in Education. Retrieved from http://cpre.org/images/stories/cpre_pdfs/AC-08.pdf

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice, 17*(4), 399-417.

Thompson, M. (2012). *Principal perceptions of experiences that helped to improve their practice as instructional leaders* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3579768).

Ulmer, M. A. (2002). *Principals' leadership practices in the context of state assessments* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. AAI3050209).

U.S. Department of Education. (1994). *Goals 2000: Educate America Act*. Washington, DC: Author. Retrieved from https://www2.ed.gov/legislation/GOALS2000/TheAct/index.html

U.S. Department of Education. (2002). *No Child Left Behind Act of 2001*. Washington, DC: Author. Retrieved from http://www.ed.gov/policy/elsec/legs/esea02/107-110.pdf

U.S. Department of Education. (2009). *Race to the Top executive summary*. Washington, DC: Author. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education (2007). *Alignment with the No Child Left Behind Act.* Washington, DC: Author. Retrieved from https://www2.ed.gov/policy/speced/guid/idea/tb-nclb-align.pdf

U.S. Department of Education (2002). *No Child Left Behind*. Washington, DC: Author. Retrieved from https://www2.ed.gov/nclb/landing.jhtml

U.S. Department of Education (2004). *The Individuals with Disabilities Education Improvement Act*. Washington, DC: Author. Retrieved from https://www2.ed.gov/policy/speced/guid/idea/idea2004.html

U.S. Department of Education (2007). *Questions and answers on Response to Intervention (RtI) and Early Intervening Services (EIS).* Washington, DC: Author. Retrieved from https://sites.ed.gov/idea/files/07-0021.RTI_.pdf

U.S. Department of Education. (2009). *Race to the Top executive summary*. Washington, DC: Author. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education (2010). *President's blueprint for reform: The reauthorization of the Elementary and Secondary Education Act.* Washington,

DC: Author. Retrieved from http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf

U.S. Department of Education (2012). *ESEA Flexibility.* Washington, DC: Author. Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

U.S. Department of Education. (2015). *Every Student Succeeds Act*. Washington, DC: Author. Retrieved from http://www.ed.gov/essa

Valentine, J. W., & Prater, M. (2011). Instructional, transformational, and managerial leadership and student achievement: High school principals make a difference. *NASSP Bulletin, 95*(1), 5-30.

van Buuren S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research, 16*(3) 219–24.

Volante, L. (2010). Assessment of, for, and as learning within schools: Implications for transforming classroom practice. *Action in Teacher Education, 31*(4), 66-75.

Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education, 34*(2), 239-255.

Volante, L., & Cherubini, L. (2011). The challenges school administrators face in building assessment literacy. *Assessment Matters, 3*, 161-182.

Wainer, H. (2010). 14 conversations about three things. *Journal of Educational and Behavioral Statistics, 35*(1), 5–25.

Waters, J. T., Marzano, R. J., & McNulty, B. A. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Aurora, CO: Mid-continent Research for Education and Learning.

Wayman, J. C., Midgley, S., & Stringfield, S. (2006). Leadership for data-based decision making: Collaborative data teams. In A. Danzig, K. Borman, B. Jones, & B. Wright (Eds.), *New models of professional development for learner centered leadership* (pp. 189-206). Mahwah, NJ: Erlbaum. Retrieved from http://www.waymandatause.com/wp-content/uploads/2013/11/Wayman-Midgley-Stringfield.pdf

Webb, N. L. (2002, December). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states.* Washington, DC: Council of Chief State School Officers.

Welch, L., Adams, G., Brown, J. L., Welch, A., & Marzano, R. J. (2008). *The art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. New York, NY: Cambridge University Press.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30,* 377–399.

William, D. & Leahy, S. (2007). A theoretical foundation for formative assessment. In J.H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 29-42). New York, NY: Teachers College Press.

Witziers, B., Bosker, R. J., & Kruger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly, 39*(3), 398-425.

Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*(4), 521–548.

Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *Education Policy Analysis Archives, 18*(19), 1-40.

Young, M. D., Crow, G., Ogawa, R., & Murphy, J. (Eds.) (2009). *The handbook of research on the education of school leaders.* New York, NY: Routledge.

Youngs, P. (2013). *Using teacher evaluation reform and professional development to support common core assessments.* Washington, DC: Center for American Progress. Retrieved from http://cdn.americanprogress.org/wp-content/uploads/2013/01/YoungsCommonCore.pdf

Carrie Elizabeth Eubank Morris

VITA

Educational Background

Educational Specialist (EdS): Curriculum and Instruction, Concentration in School Psychology	University of South Florida, Tampa, FL	May 2004
Masters of Arts (MA): School Psychology	University of South Florida, Tampa, FL	Aug 2002
Bachelors of Arts (BA): Psychology	University of Louisville, Kentucky	Dec 1999

Professional Positions

Associate Dean, Ultimate Medical Academy	Tampa, FL	Jan 2016 – present
General Education Chair, Ultimate Medical Academy	Tampa, FL	Jan 2016 – Jun 2016
Senior Supervisor, Student Services Pasco County Schools	Land O'Lakes, FL	Aug 2013 – Jan 2016
Educational Consultant Pearson Education, Inc.	New York, NY	Jul 2010 – Aug 2013
Independent Consultant, Reading and Language Arts Centers, Inc.	Detroit, MI	Jul 2010 – Aug 2013
District Data Coordinator, Woodford County Schools	Versailles, KY	Jul 2010 – Aug 2013
School Psychologist, Woodford County Schools	Versailles, KY	Jul 2007 – Aug 2013
School Psychologist, Pasco County Schools	Land O'Lakes, FL	Jul 2003 – Jun 2007

Professional Memberships

Online Learning Consortium (OLC)	2017 – present
Consortium for Research on Educational Assessment and Teaching Effectiveness (CREATE)	2017 – present
Association for Supervision and Curriculum Development (ASCD)	2016 – present
National Association of School Psychologists (NASP)	2001 – present

Professional Certifications

School Psychologist (PK-12), Florida Department of Education